

Kam, Chih-Min (AU1653)

To: Schreiber, David
Cc: Kam, Chih-Min (AU1653)
Subject: sequence search for 09/913,351

David:

Please do amino acid search with commercial data file. Thanks

Chih-Min Kam

NAME: Chih-Min Kam
AU: 1653
Date: 9/8/02
Room: 10D16
Mail Box: CM1, 9B01
Tel: 308-9437

Application No: 09/640,957.

Please search the following amino acid sequence(s) with commercial data file.

The claims cite a dimer (a conjugate) which has a minor subunit and a major subunit, where the minor subunit is

1. SEQ ID NO:2
2. SEQ ID NO:4

and the major subunit is

3. SEQ ID NO:3
4. SEQ ID NO:5

What is the best way to search this?

SEQ ID NO: 1 variable aa

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:39:28 : Search time 94.6 seconds
(without alignments)
43.443 Million cell updates/sec

Title: US-09-913-351-2
Perfect score: 186
Sequence: 1 ESKGREGSSQCRQEVQRKDLSSCERYLRQSSRR 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186	100.0	37	21	AA15420
2	61.5	33.1	110	15	AA62762
3	61.5	33.1	140	10	AA91891
4	58	31.2	148	21	AA58049
5	58	31.2	148	21	AA58050
6	57.5	30.9	111	19	AA86264
7	57.5	30.9	111	19	AA86265
8	57	30.6	904	22	AB62853
9	55	29.6	156	20	AA40973
10	55	29.6	157	18	AAW24164
11	55	29.6	157	18	AAW24153

12	55	29.6	157	20	AA15245	Peanut allergen, A
13	55	29.6	157	22	AAU04707	Anaphylactic anti
14	55	29.6	166	20	AA140958	Recombinant Ara h
15	55	29.6	207	22	AA82383	Peanut allergen Ar
16	54.5	29.3	165	22	ABG20170	Novel human diago
17	54	29.0	430	21	AA144281	Murine A20 Binding
18	53.5	28.8	34	21	AA15421	Castor bean 2S alb
19	53	28.5	593	22	AB858343	Drosophila melanog
20	53	28.5	917	21	AA539320	A Bcl-2 associated
21	53	28.5	930	22	ABG16236	Novel human diago
22	53	28.5	1450	22	ABG04173	Novel human diago
23	53	28.5	1462	22	ABG07330	Novel human diago
24	53	28.5	1462	22	AAW2582	Human protein sequ
25	53	28.5	1475	22	AB11449	Human P13-kinase h
26	52.5	28.2	158	18	AAW23586	Mablinin MBLI from
27	52.5	28.2	158	18	AAW23588	Mablinin MBLIIT fr
28	52.5	28.2	758	22	ABG05454	Novel human diago
29	52.5	28.2	758	22	ABG07092	Novel human diago
30	52.5	28.2	758	22	ABG10119	Novel human diago
31	52.5	28.2	758	22	ABG10239	Novel human diago
32	52.5	28.2	758	22	ABG10412	Novel human diago
33	52.5	28.2	758	22	ABG14939	Novel human diago
34	52.5	28.2	797	22	ABG19906	Novel human diago
35	52	28.0	1775	22	AB864008	Drosophila melanog
36	52	28.0	2061	22	AB866938	Drosophila melanog
37	51.5	27.7	91	22	AA673889	Human colon cancer
38	51.5	27.7	157	19	AAW53261	Amino acid sequenc
39	51.5	27.7	171	15	AAW53579	Synthetic 2S seed
40	51.5	27.7	174	22	AA72901	Flax 2S storage pr
41	51	27.4	52	22	AB829553	Peptide #2204 enco
42	51	27.4	52	22	AB832337	Peptide #4988 enco
43	51	27.4	52	22	AB834736	Peptide #2742 enco
44	51	27.4	52	22	AB837585	Peptide #5101 enco
45	51	27.4	52	22	AB822889	Protein #488 enco

ALIGNMENTS

RESULT	1	
AA15420	AA15420 standard; peptide; 37 AA.	
XX		
XX	AA15420;	
XX		
DT	17-JAN-2001 (first entry)	
XX		
DE	Castor bean 2S albumin storage protein amino acids 36-72.	
XX		
KW	Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;	
KW	polysaccharide; mannose; galactose; castor bean; immune response; human;	
KW	2S albumin storage protein; animal; tumour necrosis factor; mononuclear;	
KW	phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.	
XX		
OS	Ricinus communis.	
PN	WO200050087-A1.	
PD	31-AUG-2000.	
PF	21-OCT-1999; 99WO-ES00338.	
XX		
PR	26-FEB-1999; 99ES-000408.	
PA	(INFA-) IND FARM CANTABRIA SA.	
XX		
PI	Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;	
PI	Pivel Ranieri JP, Gimenez Gallego G, Matji Tuduri JA;	
XX		
DR	WPI; 2000-558369/51.	
XX		
PT	New glycoconjugate, useful for treating immunological disorders,	
PT	comprises polysaccharide from Candida utilis and polypeptide from	


```

PN WO9724139-A1.
XX
XX 10-JUL-1997.
PD
XX 23-SEP-1996; 96WO-US15222.
PF
XX 04-MAR-1996; 96US-0610424.
PR 29-DEC-1995; 95US-0009455.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
PI Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
XX
XX WPI: 1997-363453/33.
DR N-PSDB; AAT76614.
XX
XX Peanut allergens Ara hi and Ara hII - used for vaccination and in
PT two-site monoclonal antibody based ELISA
PS
XX Claim 31; Page 219; 354pp; English.
XX
CC This polypeptide comprises major peanut allergen Ara hII.
CC Its sequence was deduced from cDNA clone p38 (AAT76614), isolated
CC from peanut seed cDNA using a primer (see AAT76617) based on an
CC isolated Ara hi peptide (see AAW24151). The sequence shows
CC significant homology with the conglutin family of seed storage
CC proteins of other legumes. The allergen is recognised by serum
CC IgE from a large proportion of individuals with peanut
CC hypersensitivity. Ara hII and Ara hi (see AAW24149-50) can be used to
CC raise monoclonal antibodies which are used in a specific two-site
CC Mab ELISA for the detection of Ara hi or Ara hII (claimed). IgE-
CC binding Ara hII antigen epitopes (see AAW24188-93) may be used in
CC vaccines to protect against allergic reactions to peanut allergens,
CC e.g. anaphylactic shock.
XX
SQ Sequence 157 AA;

```

```

Query Match 29.6%; Score 55; DB 18; Length 157;
Best Local Similarity 30.6%; Pred. No. 10;
Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;

```

```

QY 1 ESKGERGSSSQOCROEVQRKDLSCCERYLRQSSSR 36
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 23 elgqdr-----rcsqqleranlrpceqhlmqkigr 52

```

```

RESULT 11
AAW24153
ID AAW24153 standard; Protein; 157 AA.
XX
XX AAW24153;
AC
XX
DT 29-DEC-1997 (first entry)
XX
DE Peanut allergen Ara hII.
XX
KW Peanut; seed storage protein; allergen; allergy; hypersensitivity;
KW vaccine; anaphylactic shock; immunotherapy; therapy;
KW monoclonal antibody; ELISA; analysis; Ara hII.
XX
OS Arachis hypogaea strain Florunner.
XX
XX WO9724139-A1.
XX
XX 10-JUL-1997.
PD
XX
XX 23-SEP-1996; 96WO-US15222.
PF
XX 04-MAR-1996; 96US-0610424.
PR 29-DEC-1995; 95US-0009455.
XX
XX (UYAR-) UNIV ARKANSAS.
PA

```

```

XX Bannon GA, Burks AW, Cockrell G, Helm RM, Stanley JS;
XX
XX WPI: 1997-363453/33.
DR N-PSDB; AAT76615.
XX
XX Peanut allergens Ara hi and Ara hII - used for vaccination and in
PT two-site monoclonal antibody based ELISA
PS
XX Claim 31; Page 198; 354pp; English.
XX

```

```

CC This polypeptide comprises major peanut allergen Ara hII.
CC Its sequence was deduced from a cDNA clone (AAT76615) isolated
CC from peanut seed cDNA using a primer (see AAT76617) based on an
CC isolated Ara hi peptide (see AAW24151). The sequence shows
CC significant homology with the conglutin family of seed storage
CC proteins of other legumes. The allergen is recognised by serum
CC IgE from a large proportion of individuals with peanut
CC hypersensitivity. Ara hII and Ara hi (see AAW24149-50) can be used to
CC raise monoclonal antibodies which are used in a specific two-site
CC Mab ELISA for the detection of Ara hi or Ara hII (claimed). IgE-
CC binding Ara hII antigen epitopes (see AAW24188-93) may be used in
CC vaccines to protect against allergic reactions to peanut allergens,
CC e.g. anaphylactic shock.
XX
SQ Sequence 157 AA;

```

```

Query Match 29.6%; Score 55; DB 18; Length 157;
Best Local Similarity 30.6%; Pred. No. 10;
Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;

```

```

QY 1 ESKGERGSSSQOCROEVQRKDLSCCERYLRQSSSR 36
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 23 elgqdr-----rcsqqleranlrpceqhlmqkigr 52

```

```

RESULT 12
AAV15245
ID AAV15245 standard; Protein; 157 AA.
XX
XX AAV15245;
AC
XX
DT 09-NOV-1999 (first entry)
XX
DE Peanut allergen, Ara h 2, amino acid sequence.
XX
KW allergy; immune response; transgenic; allergen; epitope;
KW immunoglobulin E; Ig E; binding site; peanut.
XX
OS Arachis hypogaea.
XX
XX WO9938978-A1.
XX
XX 05-AUG-1999.
PD
XX
XX 29-JAN-1999; 99WO-US02031.
PF
XX 27-AUG-1998; 98US-0141220.
PR 31-JAN-1998; 98US-0073283.
PR 13-FEB-1998; 98US-0074590.
PR 13-FEB-1998; 98US-0074624.
PR 13-FEB-1998; 98US-0074633.
XX
XX (SOSI/) SOSIN H.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX (UYNY ) UNIV NEW YORK MT SINAI SCHOOL MEDICINE.
XX
XX Bannon GA, Burks AW, Sampson HA, Sosin H;
XX
XX WPI: 1999-479189/40.
DR N-PSDB; AAZ06383.
XX

```


Query Match 29.6%; Score 55; DB 20; Length 166;
 Best Local Similarity 30.6%; Pred. No. 11;
 Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;
 Oy 1 ESKGERBSSSQOCROEVQRKDLSCERYLRQSSSR 36
 Db 21 elgqdr-----rcsqqlerantlrpceqhlmqkxqr 50

Search completed: September 9, 2002, 12:42:39
 Job time: 191 sec

RESULT 15
 AAB82383
 ID AAB82383 standard; Protein; 207 AA.
 XX
 AC AAB82383;
 XX
 DT 23-JUL-2001 (first entry)
 XX
 DE Peanut allergen Ara h2 gene product.
 XX
 KW Peanut; allergen; Ara h2; transgenic plant; allergy.
 XX
 OS Arachis hypogea.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= signal_peptide
 FT 22..207
 FT Protein /label= Mature_protein
 XX
 PN W0200136621-A2.
 XX
 PD 25-MAY-2001.
 XX
 PE 20-NOV-2000; 2000WO-US31657.
 XX
 PR 19-NOV-1999; 99US-0167255.
 XX
 PA (UYAL-) UNTV ALABAMA A & M.
 XX
 PI Dodo HW, Arntzen CJ, Konan KN, Viquez OM;
 XX
 DR WPI; 2001-355630/37.
 DR N-PSDB; AAF90336, AAF90337.
 XX
 PT Producing transgenic peanut plants that produce allergen-free seeds,
 PT useful in non-allergenic foods, by antisense or sense co-suppression of
 PT allergen-encoding genes -
 XX
 PS Example 1; Fig 2; 72pp; English.
 XX
 CC The present sequence is that of the peanut allergenic protein (AP)
 CC encoded by the Ara h2 gene (see AAF90336). The invention relates
 CC to a method for producing a peanut plant having reduced, or
 CC undetectable, AP content in its seed. A peanut plant cell is
 CC transformed with a DNA construct containing an antisense AP gene
 CC and/or sense AP gene, or their fragments, regenerated to plants,
 CC and fertile transgenic plants that produce seeds with reduced AP
 CC content are identified. The AP sense or antisense gene may
 CC comprise at least a portion of the Ara h2 gene sequence. The seeds
 CC are useful for preparation of allergen-free foods. Recombinant AP
 CC may be produced and used to produce antibodies useful for detecting
 CC AP in foods, and for treatment or prevention of peanut allergy.
 CC
 XX
 SQ Sequence 207 AA;

Query Match 29.6%; Score 55; DB 22; Length 207;
 Best Local Similarity 30.6%; Pred. No. 14;
 Matches 11; Conservative 10; Mismatches 9; Indels 6; Gaps 1;

Oy 1 ESKGERBSSSQOCROEVQRKDLSCERYLRQSSSR 36
 Db 26 elgqdr-----rcsqqlerantlrpceqhlmqkxqr 55

Db 33 DNDDENOPLCRFOFOOHHRACQYIRRRRAQR 65

RESULT 2

US-08-670-186-6

; Sequence 6, Application US/08670186

Patent No. 5859343

GENERAL INFORMATION:

APPLICANT: SUN, SAMUEL S.M.

APPLICANT: XIONG, LIWEN

APPLICANT: HU, ZHONG

APPLICANT: CHEN, HANG

TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FORSTER

STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/670,186

FILING DATE: 21-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 23461-20007.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 822-0168

TELEX: 90-4030 MRSNFOERSWSH

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 158 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-670-186-6

Query Match 28.2%; Score 52.5; DB 2; Length 158;

Best Local Similarity 27.3%; Pred. No. 5.1;

Matches 9; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

QY 5 EREGSSSQOCROEVOR-KDLSSCERLRRSSSR 36

Db 33 DNDDENOPLCRFOFOOHHRACQYIRRRRAQR 65

RESULT 3

US-08-706-391B-7

; Sequence 7, Application US/08706391B

Patent No. 6174725

GENERAL INFORMATION:

APPLICANT: ANDERSON, OLIN D

TITLE OF INVENTION: ALTERING DOUGH VISCOELASTICITY WITH MODIFIED GLUTENINS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: USDA, AGRICULTURAL RESEARCH SERVICE, PACIFIC WEST

STREET: 800 BUCHANAN STREET

CITY: ALBANY

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94710

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/706,391B

FILING DATE: 30-Aug-1996

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CONNOR, MARGARET A

REGISTRATION NUMBER: 30,043

REFERENCE/DOCKET NUMBER: 0235.95/USDA96-001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 559-6067

TELEFAX: (510) 559-5777

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 157 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-08-706-391B-7

Query Match 27.7%; Score 51.5; DB 4; Length 157;

Best Local Similarity 35.5%; Pred. No. 6.9;

Matches 11; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

QY 5 EREGSSSQOC--ROEVOR-KDLSSCERLRRSSSR 32

Db 2 EGRASELOCDRELJOEHELKACQOVMDQ 32

RESULT 4

US-08-911-853-9

; Sequence 9, Application US/08911853

Patent No. 6048710

GENERAL INFORMATION:

APPLICANT: Geritise, Gijstbert

APPLICANT: Quax, Wilhelmus J.

TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED

TITLE OF INVENTION: EXPRESSION LEVELS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International

STREET: 925 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1013

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/911,853

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/699,092

FILING DATE: 16-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Glaister, Debra J

REGISTRATION NUMBER: 33,888

REFERENCE/DOCKET NUMBER: GC361-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-846-7620

TELEFAX: 650-845-6504

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

Query Match 26.3%; Score 49; DB 3; Length 873;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 GREGSSSQCRQEVQRDLSCERYL 30
DB 362 GEKFKQIQRFHQGVQLIDFSCERYL 388

RESULT 8

US-09-546-238-2
Sequence 2, Application US/09546238
Patent No. 6316225
GENERAL INFORMATION:
APPLICANT: Olsen, Henrik S.
APPLICANT: Ruben, Steven M.
APPLICANT: Sonenberg, Nahum
APPLICANT: Method, Nathalie
APPLICANT: Rom, Eran
TITLE OF INVENTION: Human Ptl-1-like Subunit Protein (hPtl) Polynucleotides
TITLE OF INVENTION: (as amended)
FILE REFERENCE: 1488.0700002
CURRENT APPLICATION NUMBER: US/09/546,238
CURRENT FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: US 60/033,151
PRIOR FILING DATE: 1996-12-13
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 873
TYPE: PRT
ORGANISM: Homo sapiens
US-09-546-238-2

Query Match 26.3%; Score 49; DB 4; Length 873;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 GREGSSSQCRQEVQRDLSCERYL 30
DB 362 GEKFKQIQRFHQGVQLIDFSCERYL 388

RESULT 9

US-08-444-231-19
Sequence 19, Application US/08444231
Patent No. 5652210
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J.
APPLICANT: SHAPIRO, JOHN P.
APPLICANT: KIEFER, MICHAEL C.
TITLE OF INVENTION: NOVEL FAS PROTEIN AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,231
FILING DATE: 18-MAY-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/152,443
FILING DATE: 15-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20006.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-231-19

Query Match 26.1%; Score 48.5; DB 1; Length 314;
Best Local Similarity 26.9%; Pred. No. 39;
Matches 7; Conservative 14; Mismatches 4; Indels 1; Gaps 1;

QY 9 SSSQCCRQEVQRDL-SSCERYLROS 33
DB 160 TSNFKCKEVRKREYQKCRKRKEN 185

RESULT 10

US-08-152-443A-19
Sequence 19, Application US/08152443A
Patent No. 5663070
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J.
APPLICANT: SHAPIRO, JOHN P.
APPLICANT: KIEFER, MICHAEL C.
TITLE OF INVENTION: NOVEL FAS PROTEIN AND METHODS OF USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,443A
FILING DATE: 15-NOV-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LEHNHARDT, SUSAN K.
REGISTRATION NUMBER: 33,943
REFERENCE/DOCKET NUMBER: 23647-20006.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-152-443A-19

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: September 9, 2002, 12:40:58 ; Search time 22.97 Seconds

(without alignments)
62.369 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186
Sequence: 1 ESKGEREGSSSQCRQEVQRKDLSCRCRYLRSSNR 37

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	186	100.0	2S5_RICCO	P01089 ricinus com
2	62	33.3	SWAP_CAEEL	Q10580 caenorhabdi
3	61.5	33.1	2S5_BEREX	P04403 bertolletti
4	59	31.7	2S55_HELAN	P15461 helianthus
5	58	31.2	PUB1_WHEAT	Q10464 triticum ae
6	56	30.1	CG2S_LUPAN	P09930 lupinus ang
7	55	29.6	GLT0_WHEAT	P10387 triticum ae
8	55	29.6	GLT3_WHEAT	P08488 triticum ae
9	55	29.6	Y218_HUMAN	Q93075 homo sapien
10	53.5	28.8	2SS4_CAPMA	P80353 capparis ma
11	53.5	28.8	2SS3_CAPMA	P08489 triticum ae
12	53.5	28.8	GLT4_WHEAT	P10388 triticum ae
13	53	28.5	GLT5_WHEAT	P10387 triticum ae
14	53	28.5	RNI_AERY	Q07465 aeromonas h
15	53	28.5	GSH1_BUCAL	P57485 buchera ap
16	53	28.5	PK3G_HUMAN	O75747 homo sapien
17	52	28.0	RL37_SCHMA	Q44125 schistosoma
18	51.5	27.7	2SS1_CAPMA	P80351 capparis ma
19	51	27.4	SBP_SOYBN	Q04672 glycine max
20	51	27.4	ORA_PLAFN	P16405 plasmodium
21	50	26.9	YC98_SCHPO	Q09882 schizosacch
22	50	26.9	VITI_FUNHE	Q90508 fundulus he
23	49.5	26.6	GL19_ORYSA	P29835 oryza sativ
24	49.5	26.6	SC2_OCTDO	P27010 octopus dof
25	49	26.3	COPE_HUMAN	O14579 homo sapien
26	49	26.3	DMP1_RAT	P8183 ratius norv
27	49	26.3	VE2_HPV12	P26782 human papil
28	49	26.3	IF39_HUMAN	P55884 homo sapien
29	48.5	26.1	HAP2_KLULA	P33768 kluyveromyc
30	48.5	26.1	HMID_DROAN	P22544 drosophila
31	48.5	26.1	CARE_MOUSE	Q99K80 mus musculu
32	48	25.8	HS1_MOUSE	P49710 mus musculu
33	48	25.8	AMAI_PLAFR	P22622 plasmodium

34	48	25.8	815	1	GTRB_MYXXA	O33367 myxococcus
35	48	25.8	1004	1	CARE_HUMAN	Q9BX16 homo sapien
36	48	25.8	1506	1	PK3G_MOUSE	O70167 mus musculu
37	48	25.8	1898	1	TRHY_HUMAN	O07283 homo sapien
38	48	25.8	4473	1	PIEL_CRIGR	O91155 cricetus
39	48	25.8	4687	1	PIEL_RAT	P30427 ratius norv
40	47.5	25.5	82	1	V467_TREPA	O83480 treponema p
41	47.5	25.5	110	1	2SS1_BRANA	P24565 brassica na
42	47.5	25.5	158	1	2SS5_SOYBN	P19594 glycine max
43	47.5	25.5	235	1	ARDH_HUMAN	P41227 homo sapien
44	47.5	25.5	487	1	MDM2_CANFA	P56950 canis famli
45	47	25.3	376	1	DUT_HSV62	P52541 human herpe

ALIGNMENTS

```

RESULT 1
ID 2S5_RICCO STANDARD: PRT; 258 AA.
AC P01089;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91105729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preprotein
RT may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
RT gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
RT protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
RT inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY 2 DISULFIDE BONDS.
CC -1- PPM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CC CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X54158; CAA38097.1; -.
DR PIR; A01328; RZCS.
DR PIR; S11499; S11499.
DR PIR; S11500; S11500.
DR PIR; S11501; S11501.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR01768; Cereals_1ryp_aml_1nh.
DR Pfam; PF00234; trypt_alpha_aml_1_2.
DR SMART; SM00499; AAI; 2.
KM Seed storage protein; Signal; Allergen.
FT SIGNAL 1 21 PROBABLE.
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT PROPEP 191 193
FT CHAIN 194 258
FT MOD_RES 194 194 2S ALBUMIN, LARGE CHAIN.
FT CONFLICT 222 222 PYROLIDONE CARBOXYLIC ACID.
FT CONFLICT 226 229 E -> Q (IN REF. 3).
FT CONFLICT 234 234 MISSING (IN REF. 3).
FT CONFLICT 255 255 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;

```

```

Query Match 100.0%; Score 186; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 2.1e-16;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ESKGEREGSSQOCROEVORRDLSSCERYLRFOSSRR 37
Db 36 ESKGEREGSSQOCROEVORRDLSSCERYLRFOSSRR 72

```

RESULT 2

```

SWAP_CAEEL STANDARD; PRT; 749 AA.

```

```

AC Q10380;

```

```

DT 01-OCT-1996 (Rel. 34, Created)

```

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 01-OCT-1996 (Rel. 34, Last annotation update)

```

```

DE SWAP protein (Suppressor of white apricot protein homolog).

```

```

GN SWAP.

```

```

OS Caenorhabditis elegans.

```

```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

```

OC Rhabditidae; Peleodermidae; Caenorhabditis.

```

```

OX NCB1_TaxID=6239;

```

```

RN [1]

```

```

RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=95061415; PubMed=7971282;

```

```

RA Spikes D.A., Kramer J., Bingham P.M., van Doren K.;

```

```

RT "SWAP pre-mRNA splicing regulators are a novel, ancient protein

```

```

RT family sharing a highly conserved sequence motif with the prp1

```

```

RT family of constitutive splicing proteins.";

```

```

RL Nucleic Acids Res. 22:4510-4519(1994).

```

```

CC -1- FUNCTION: IT IS A REGULATOR OF PRE-MRNA SPLICING (AND, POSSIBLY,

```

```

CC OF OTHER RNA PROCESSING EVENTS). IT MAY REGULATE ITS OWN

```

```

CC EXPRESSION AT THE LEVEL OF RNA PROCESSING.

```

```

CC -1- SIMILARITY: TO DROSOPHILA SWA.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/

```

```

CC or send an email to license@lsb-sib.ch).

```

```

CC EMBL; U06933; AAA64937.1; -.

```

```

CC EMBL; U06933; AAA64938.1; -.

```

```

DR InterPro; IPR000061; Surp.

```

```

DR Pfam; PF01805; Surp. 2.

```

```

KW Transcription regulation; RNA-binding; mRNA splicing; Repeat.

```

```

FT DOMAIN 8 124 DRY CEEERYL.
FT DOMAIN 166 431 2 X REPEATS OF THE SURP MOTIF.
FT REPEAT 166 209 SURP MOTIF 1.
FT REPEAT 391 431 SURP MOTIF 2.
FT DOMAIN 688 749 ARG/LYS/SER-RICH (HIGHLY BASIC).
SQ SEQUENCE 749 AA; 86309 MW; 7F2627E37CD1C590 CRC64;

```

```

Query Match 33.3%; Score 62; DB 1; Length 749;
Best Local Similarity 46.2%; Pred. No. 1.9;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

```

```

OY 5 EREGSSQOCROEVORRDLSSCERYL 30
Db 71 KRNSPSEOCPTFAMEDMCEERYL 96

```

RESULT 3

```

ID 25S_BEREX STANDARD; PRT; 146 AA.

```

```

AC P04403; P04402;

```

```

DT 20-MAR-1987 (Rel. 04, Created)

```

```

DT 01-OCT-1989 (Rel. 12, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE 2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).

```

```

GN BEZS1 AND BEZS2.

```

```

OS Bertholletia excelsa (Brazil nut).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

OC Asteridae; Ericales; Lecythidaceae; Bertholletia.

```

```

OX NCB1_TaxID=3645;

```

```

RN [1]

```

```

RP SEQUENCE FROM N.A.

```

```

RA Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;

```

```

RT "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein

```

```

RT exceptionally rich in methionine.";

```

```

RL Plant Mol. Biol. 8:239-250(1987).

```

```

RN [2]

```

```

RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=91370890; PubMed=1840683;

```

```

RA Gander E.S., Holmstroem K.O., de Paiva G.R., de Castro L.A.B.,

```

```

RA Carneiro M., Grossi de Sa M.F.;

```

```

RT "Isolation, characterization and expression of a gene coding for a 2S

```

```

RT albumin from Bertholletia excelsa (Brazil nut)."

```

```

RL Plant Mol. Biol. 16:437-448(1991).

```

```

RN [4]

```

```

RP SEQUENCE OF 37-64 AND 70-142.

```

```

RX MEDLINE=87004679; PubMed=3758080;

```

```

RA Ampe C., van Damme J., de Castro L.A.B., Sampaio M.J.A.M.,

```

```

RA van Montagu M., Vandekerckhove J.;

```

```

RT "The amino-acid sequence of the 2S sulphur-rich proteins from seeds

```

```

RT of Brazil nut (Bertholletia excelsa H.B.K.).";

```

```

RL Eur. J. Biochem. 159:597-604(1986).

```

```

CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.

```

```

CC -1- SUBUNIT: THE NATIVE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN

```

```

CC LINKED BY DISULFIDE BONDS.

```

```

CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/

```

```

CC or send an email to license@lsb-sib.ch).

```

```

CC EMBL; M17146; AAA33010.1; -.

```

```

DR EMBL; X57027; CAA40343.1; -.

```

```

DR EMBL; X57028; CAA40344.1; -.

```



```

FT PROPEP 20 29
FT CHAIN 30 148
FT DOMAIN 68 73
SO SEQUENCE 148 AA; 16792 MW; 327904B4EBEC2C16 CRC64;

Query Match
Best Local Similarity 31.28; Score 58; DB 1; Length 148;
Matches 13; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 8 GSSGQCRQEVQRKDLSSCERYLRQ 32
| ||||| | | ||||| : : :
Db 34 GGSQCCPCPE--RPKSSCKDYWE 56

RESULT 6
CG2S_LUPAN STANDARD; PRT; 37 AA.
AC P09350;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Conglutin delta-2 small chain.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
ON NCBI_TaxID=3871;
RX [1]
RP SEQUENCE.
RC STRAIN-CV. WHITE;
RA Lilly G.G., Inglis A.S.;
RT "Amino acid sequence of conglutin delta, a sulfur-rich seed protein
RT of Lupinus angustifolius L. Sequence homology with the C-III alpha-
RT amylase inhibitor from wheat."
FBBS Lett. 195:235-241(1986).
CL -1- SUBUNIT: DIMER OF A SMALL CHAIN AND A LARGE CHAIN LINKED BY TWO
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR; A23617; A23617.
KW Seed.
FT DISUFID 8 8 INTERCHAIN (WITH C-29 OF LARGE CHAIN).
FT DISUFID 20 20 INTERCHAIN (WITH C-17 OR C-18 OF LARGE
FT FT CHAIN).
FT DOMAIN 29 37 GLU/GLN-RICH.
SQ SEQUENCE 37 AA; 4598 MW; 89784D55A5A1493A CRC64;

Query Match
Best Local Similarity 30.18; Score 56; DB 1; Length 37;
Matches 9; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 9 SSSGQCRQEVQRKDLSSCERYLRQ 32
| | | | | : : : | | : |
Db 3 SSSGCRQLQVLRHCENHIIDQ 26

RESULT 7
GLT3_WHEAT STANDARD; PRT; 648 AA.
ID GLT3_WHEAT
AC P10387;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit D10 precursor.
GN GLU-D1-2b.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-CV. CHEYENNE;
RX MEDLINE=89098419; PubMed=2563152;
RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA Malpica-Romero J.M.;
RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT Cheyenne."
RL Nucleic Acids Res. 17:461-462(1989).
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISCIO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GOOPGQ AND
CC GOOPGQGOOGYPTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb.ch/announce/
CC or send an email to license@isb-sid.ch.)
CC -----
CC EMBL; X12929; CAA31396.1; -
CC DR PIR; S04832; S04832.
CC DR HSSP; P01088; 1BFA.
CC DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
CC DR InterPro; IPR001419; Glutenin.
CC DR Pfam; PF00234; tryp_alpha_amyl_1.
CC DR PRINTS; PR00210; GLUTENIN.
CC KW Seed storage protein; Repeat; Multigene family; Signal.
FT FT SIGNAL 1 21
FT CHAIN 22 648 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT FT D10.
FT DOMAIN 147 610 REPEATS.
SQ SEQUENCE 648 AA; 69629 MW; FE98FID44B9E9AF1 CRC64;

Query Match
Best Local Similarity 29.68; Score 55; DB 1; Length 648;
Matches 11; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

OY 5 EREGSSQCRQEVQRKDLSSCERYLRSSRR 37
| | | | | : : : | | : |
Db 22 EGEASRQLQCRERLESSLACRQVVDQLAGR 54

RESULT 8
GLT3_WHEAT STANDARD; PRT; 660 AA.
ID GLT3_WHEAT
AC P08458;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit 12 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. CHINESE SPRING;
RX MEDLINE=86041882; PubMed=3840586;
RA Thompson R.D., Bartels D., Harberd N.P.;
RT "Nucleotide sequence of a gene from chromosome 1D of wheat encoding a
RT HMW-glutenin subunit."
RL Nucleic Acids Res. 13:6833-6846(1985).
CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE

```



```

CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -1- SUBUNIT. DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND
CC GQPGQGGGQYPPS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03041; CAA26847.1; -.
DR PIR: A24266; A24266.
DR HSSP: P01088; 1BFA.
DR InterPro: IPR001768; Cereal_tryp-amy1_inh.
DR InterPro: IPR001419; Glutenin.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00210; GLUTENIN.
DR Seed storage protein; Repeat; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 660
FT DOMAIN 125 615 REPEATS.
FT SQUENCE 660 AA; 70868 MW; 2BFD09D8C8FCCEFF CRC64;
SQ
QY Query Match 29.6%; Score 55; DB 1; Length 660;
Best Local Similarity 33.3%; Pred. No. 13;
Matches 11; Conservative 6; Mismatches 16; Indels 0; Gaps 0;
QY 5 EREGSSSQCRQEVORRKLSCERCYLRQSSRR 37
| | | | | | | | | | | | | | | | | | | | |
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 22 EGASRQLQGERELQESSLEACRGVYDQQLAGR 54
RESULT 9
Y218_HUMAN STANDARD; PRT; 761 AA.
ID Y218_HUMAN
AC 093075;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative deoxyribonuclease KIAA0218 (EC 3.1.21.-).
GN KIAA0218.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Bone marrow;
RX MEDLINE=97191544; Pubmed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayasi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
CC -1- SIMILARITY: BELONGS TO THE TAD DNASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: D86972; BAA13208.1; -.
DR InterPro: IPR001130; Tadt.
DR Pfam: PF01026; Tadt_DNase; 1.
DR PROSITE: PS01137; TADT_1; 1.
DR PROSITE: PS01090; TADT_2; 1.
DR PROSITE: PS01091; TADT_3; FALSE NEG.
KW Hypothetical protein; Hydrolase; Nuclease.
SQ SEQUENCE 761 AA; 85038 MW; 7B0504FBDD262CA8E CRC64;
QY Query Match 29.6%; Score 55; DB 1; Length 761;
Best Local Similarity 35.1%; Pred. No. 15;
Matches 13; Conservative 5; Mismatches 19; Indels 0; Gaps 0;
QY 1 ESKGERGSSSQCRQEVORRKLSCERCYLRQSSRR 37
| | | | | | | | | | | | | | | | | | | | |
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 40 QRSASRGSGSPSRKLRKAKEDDVACSRRLSWGSSRR 76
RESULT 10
2SS4_CAPMA STANDARD; PRT; 100 AA.
ID 2SS4_CAPMA
AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis masaiikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Capparis.
OX NCBI_Taxid=13395;
RN [1]
RP SEQUENCE.
RC TISSUE-Seed;
RX MEDLINE=94333405; Pubmed=8055976;
RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDICING ACTIVITY.
CC -1- SUBUNIT: HETERO DIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC InterPro: IPR003612; AAI.
CC InterPro: IPR000617; Napin.
CC Pfam: PF01631; Seedstore_2S; 1.
CC PRINTS: PR00496; NAPIN.
CC PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
DR Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28
FT NON CONS 28 29
FT CHAIN 29 100
FT DISULFID 4 49
FT DISULFID 17 38
FT DISULFID 39 87
FT DISULFID 51 95
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CRC64;
QY Query Match 28.8%; Score 53.5; DB 1; Length 100;
Best Local Similarity 41.7%; Pred. No. 2.9;
Matches 10; Conservative 8; Mismatches 5; Indels 1; Gaps 1;
QY 14 CROEVOR-KDLSCEYLRQSSRR 36
| | | | | | | | | | | | | | | | | | | | |
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 4 CRRQFOOHQHLRACQRYLRRAQR 27
RESULT 11

```

2S53_CAPMA STANDARD: PRT: 104 AA.

AC P80352:

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Mabinlin II, A and B chains (MAB III) (Sweet protein).

OS Capparis msaikai (Mabinlang).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eucosids II: Brassicales: Brassicaceae: Capparids.

OC NCBI_TaxID=13395;

XX [1]

RN NCBI_TaxID=13395;

RP SEQUENCE.

RC TISSUE=Seed:

RX MEDLINE=94333405; PubMed=8055976;

RA Nirasawa S., Nishino T., Katahira M., Uesugi S., Hu Z., Kurihara Y.;

RT Structures of heat-stable and unstable homologues of the sweet protein mabinlin. The difference in the heat stability is due to replacement of a single amino acid residue.*;

RT Eur. J. Biochem. 223:989-995(1994).

RL -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING ACTIVITY.

CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY DISULFIDE BONDS.

CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

DR Interpro: IPR003612; AAI.

DR Interpro: IPR000617; Napin.

DR Pfam: PF01631; Seedstore_25; 1.

DR PRINTS: PR00496; NAPIN.

DR ProDom: PD002498; Napin; 1.

DR SMART: SM00499; AAI; 1.

KW Seed storage protein: Albumin; Sweet-taste.

FT CHAIN 1 32 MABINLIN III, A CHAIN.

FT NON-CONS 32 33

FT CHAIN 33 104 MABINLIN III, B CHAIN.

FT DISULFID 4 53

FT DISULFID 17 42

FT DISULFID 43 91

FT DISULFID 55 99

SO SEQUENCE 104 AA; 12284 MW; 102BE3F5F24AD3D0 CRC64;

Query Match 28.8%; Score 53.5; DB 1; Length 104;

Best Local Similarity 41.7%; Pred. No. 3;

Matches 10; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

QY 14 CROEVOR-KDLSCERYLRQSSR 36

DB 4 CROFOOHOLRACORRYLRRAQR 27.

RESULT 12

GLT4_WHEAT STANDARD: PRT: 838 AA.

ID GLT4_WHEAT

AC P08489;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glutelin, high molecular weight subunit FW212 precursor.

OS Triticum aestivum (Wheat).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; Pooidae; Triticaceae; Triticum.

OC NCBI_TaxID=4565;

XX [1]

RN NCBI_TaxID=4565;

RP SEQUENCE FROM N.A.

RC STRAIN=CV. YAMHILL;

RA MEDLINE=86093674; PubMed=3001648;

RA Sugiyama T., Rafalski A., Peterson D., Soll D.G.;

RT "A wheat HMW glutenin subunit gene reveals a highly repeated Nucleic Acids Res. 13:8729-8737(1985).

CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.

CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.

CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE GROUP 1 CHROMOSOMES OF WHEAT.

CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND GQPGQGQGGQYTPS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb.ch).

CC -----

DR EMBL: X03346; CAA27052.1; -.

DR PIR: A24107; EEMTHW.

DR Interpro: IPR001419; Glutenin.

DR PRINTS: PR00210; GLUTENIN.

KW Seed storage protein; Repeat; Multigene family; Signal.

FT SIGNAL 1 21

FT CHAIN 22 838

FT DOMAIN 130 799

FT SEQUENCE 838 AA; 89174 MW; 71D715B7BDF0722D CRC64;

Query Match 28.8%; Score 53.5; DB 1; Length 838;

Best Local Similarity 35.5%; Pred. No. 25;

Matches 11; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

QY 5 ERBGSSQOCR--QEVQRKDLSCERYLRQ 32

DB 22 EGEASELOQCELELOELQERELKACQVMDQ 52

RESULT 13

GLT5_WHEAT STANDARD: PRT: 839 AA.

ID GLT5_WHEAT

AC P10388;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Glutelin, high molecular weight subunit DX5 precursor.

OS Triticum aestivum (Wheat).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae; Pooidae; Triticaceae; Triticum.

OC NCBI_TaxID=4565;

XX [1]

RN NCBI_TaxID=4565;

RP SEQUENCE FROM N.A.

RC STRAIN=CV. CHEYENNE;

RX MEDLINE=89098419; PubMed=2563152;

RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R., Malpica-Romero J.M.;

RT "Nucleotide sequences of the two high-molecular-weight glutenin genes from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv Cheyenne.*;

RT Nucleic Acids Res. 17:461-462(1989).

CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.

CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE

```
DB      EMBL; X67054; CAA47438.1; -.  
DR      PIR; A47118; A47118.  
DR      PIR; S23227; S23227.  
DR      InterPro; IPR001568; RNase_T2.  
DR      Pfam; PF00445; ribonuclease_T2; 1.  
DR      PROSITE; PS00530; RNASE_T2_1; 1.  
DR      PROSITE; PS00531; RNASE_T2_2; 1.  
KW      Hydrolase; Nuclease; Endonuclease; Periplasmic; Signal.  
FT      SIGNAL  
FT CHAIN          1           22  
FT ACT_SITE     62         62    RIBONUCLEASE.  
FT ACT_SITE     102        102   BY SIMILARITY.  
FT ACT_SITE     106        106   BY SIMILARITY.  
FT DISULFID     76         109   BY SIMILARITY.  
SQ      SEQUENCE       215 AA; 24410 MW; 4442BCE5D0F67203 CRC64;  
  
Query Match              28.5%; Score 53; DB 1; Length 215;  
Best Local Similarity    38.1%; Pred. NO. 7.3;  
Matches                  8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;  
  
OY      6 REGSSSQQCRQEVQRKDLSC 26  
      :|||: | ||::| | |  
Db      184 KEGTDAVTCSDQMOKRELPSG 204  
  
RESULT 15  
GS_H1_BUGAL  
ID GS_H1_BUGAL STANDARD: PRI: 518 AA.  
AC P57485;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glutamate--cysteine ligase [EC 6.3.2.2] (Gamma-glutamyl)cysteine synthetase) (Gamma-ECS) (GCS).  
GN GSMA OR BU407.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI_TaxId=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TOKYO 1998;  
RX MEDLINE=20445173; PubMed=10993077;  
RA Shigenobu S., Matsumae H., Hattori M., Sakaki Y., Ishikawa H.; "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Aph." Nature 407:81-86(2000).  
RL Natural 407:81-86(2000).  
CC -I- CATALYTIC ACTIVITY: ATP + L-glutamate + L-cysteine = ADP + phosphate + gamma-L-glutamyl-L-cysteine.  
CC -I- PATHWAY: FIRST STEP IN GLUTATHIONE BIOSYNTHESIS.  
CC -I- SIMILARITY: STRONG, TO E.COLI GS_H1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/about/news/or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AF001119; BAB13108.1; -  
KW Glutathione biosynthesis; Ligase; Complete proteome.  
SQ SEQUENCE       518 AA; 60979 MW; AC6659FE2C50BD4 CRC64;  
  
Query Match              28.5%; Score 53; DB 1; Length 518;  
Best Local Similarity    40.7%; Pred. No. 18;  
Matches                 11; Conservative 7; Mismatches 9; Indels 0; Gaps 0;  
  
OY      5 ERGGSSSQCRQEVQRKDLSCERYLR 31  
      |:| | | ::||:| | | |:  
Db      490 EOEFVRSHQKRRETREREDLIISPEEYR 516
```

Search completed: September 9, 2002, 12:46:06
Job time: 308 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:40:43 : Search time 77.42 Seconds
(without alignments)
82.676 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186

Sequence: 1 ESKGEREGSSQOCROEVQRKDLSCERYLRSSNR 37

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_unclassified:*

14: sp_virulent:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	41	10	Q9S874 ricinus com
2	79	42.5	153	10	Q9AUD1 sesamum ind
3	75.5	40.6	141	10	Q39649 cucurbita s
4	74	39.8	323	10	Q39928 helianthus
5	62	33.3	747	5	Q17484 caenorhabdi
6	61.5	33.1	146	10	Q9LRC2 ricini bertiollet
7	61.5	33.1	160	10	Q9SQR1 arachis hyp
8	61	32.8	572	10	Q9A064 elaeis guin
9	60.5	32.5	139	10	Q9JG58 tt virus. p
10	59	31.7	471	2	Q9EXV9 juglans reg
11	59	31.7	765	12	Q9JG55 tt virus. p
12	58	31.2	462	10	Q94I09 triticum ae
13	58	31.2	743	10	Q94I10 secale cere
14	58	31.2	754	10	Q94I12 secale cere
15	58	31.2	754	10	Q94I19 secale cere
16	58	31.2	754	10	Q94I19 secale cere

17	58	31.2	754	10	Q93WF0 secale cere
18	58	31.2	766	10	Q9SDM3 triticum ae
19	58	31.2	766	10	Q94I13 secale cere
20	58	31.2	811	10	Q94I17 triticum ae
21	57	30.6	153	10	Q99235 lupinus ang
22	57	30.6	392	5	Q95R97 arabidopsis
23	57	30.6	699	10	Q9ZU69 arabidopsis
24	57	30.6	904	5	Q9VC33 drosophila
25	56	30.1	141	10	Q9LE34 hordeum vul
26	56	30.1	141	10	Q9M3R3 hordeum vul
27	56	30.1	142	10	Q9M3V6 hordeum vul
28	56	30.1	142	10	Q9M3V5 hordeum vul
29	56	30.1	142	10	Q9M3V4 hordeum vul
30	56	30.1	147	10	Q9FS19 hordeum vul
31	56	30.1	147	10	Q9LEH8 hordeum vul
32	56	30.1	765	12	Q9JG52 tt virus. p
33	55	29.6	156	10	Q9M4D9 arachis hyp
34	55	29.6	156	10	Q94I10 arachis hyp
35	55	29.6	179	10	Q9M5N3 aegilops ve
36	55	29.6	241	5	Q23533 caenorhabdi
37	55	29.6	648	10	Q38767 aegilops ta
38	55	29.6	654	10	Q93XR5 aegilops cy
39	55	29.6	705	10	Q03871 triticum ae
40	55	29.6	707	10	Q94I16 secale cere
41	55	29.6	713	10	Q9SDM2 triticum ae
42	55	29.6	713	10	Q94I11 secale cere
43	55	29.6	713	10	Q94I18 triticum ae
44	55	29.6	713	10	Q93WM1 secale cere
45	55	29.6	720	10	Q94I16 triticum ae

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	41 AA.
Q9S874	Q9S874		
ID	01-MAY-2000 (TREMREL. 13, Created)		
AC	01-MAY-2000 (TREMREL. 13, Last sequence update)		
DT	01-JUN-2000 (TREMREL. 14, Last annotation update)		
DE	4.7 KDA NAPIN-LIKE PROTEIN SMALL CHAIN (FRAGMENT).		
OS	Ricinus communis (Castor bean).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid 1; Malpighiales; Euphorbiaceae; Ricinus.		
OX	NCBI_TaxID=3988;		
RM	[1]		
RP	SEQUENCE.		
RX	MEDLINE=97135090; PubMed=8980648;		
RA	Neumann G.M., Condron R., Polya G.M.;		
RT	"Purification and sequencing of napin-like protein small and large		
RT	chains from Momordica charantia and Ricinus communis seeds and		
RT	determination of sites phosphorylated by plant Ca(2+)-dependent		
RT	protein kinase."; Acta 1298:223-240(1996).		
RL	Biochim. Biophys. Acta 1298:223-240(1996).		
SO	SEQUENCE 41 AA; 4732 MW; 201FF2938F33173 CRC64;		
Query Match	100.0%; Score 186; DB 10; Length 41;		
Best Local Similarity	100.0%; Pred. No. 6.3e-19;		
Matches	37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 ESKGEREGSSQOCROEVQRKDLSCERYLRSSNR 37		
DB	1 ESKGEREGSSQOCROEVQRKDLSCERYLRSSNR 37		
RESULT 2	PRELIMINARY:	PRT:	153 AA.
Q9AUD1	Q9AUD1		
ID	01-JUN-2001 (TREMREL. 17, Created)		

RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RL investigating biology. The C. elegans Sequencing Consortium.";
 RN Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Tatch A.;
 RT "The sequence of C. elegans cosmid B0336.";
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U32305; AAK18860.1; -
 DR InterPro: IPR000061; Surp.
 DR Pfam: PF01805; Surp; 2.
 DR SEQUENCE 747 AA; 86068 MW; EB0A50BEF659B99E CRC64;

Query Match 33.1%; Score 62; DB 5; Length 747;
 Best Local Similarity 46.2%; Pred. No. 2.1;
 Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 5 EREGSSQCCROEVQRKDLSSCERYL 30
 Db 71 KRNGSPSECCPTAMEEDMCEERYL 96

RESULT 6
 Q9LRC2 PRELIMINARY; PRT; 146 AA.
 AC Q9LRC2;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 2S ALBUMIN
 OS Bertholletia excelisa (Brazil nut).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Ericales; Lecythidaceae; Bertholletia.
 OX NCBI_TaxID=3645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamauchi D.;
 RT "Brazil nut 2S albumin was synthesized in a transgenic French bean
 RT seed with a promoter of the gene for canavain, 7S globulin from
 RL Canavalia gladiata.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB044391; BAA96554.1; -
 DR HSSP; P01087; IB1U.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_aml1_inh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_aml1; 1.
 DR PRINTS; PR00808; AMLASEINHBR.
 DR PRINTS; PR00496; NAPIN.
 DR SMART; SM00499; AAI; 1.
 DR SEQUENCE 146 AA; 16910 MW; 4A69A196E6EC7096 CRC64;

Query Match 33.1%; Score 61.5; DB 10; Length 146;
 Best Local Similarity 40.7%; Pred. No. 0.49;
 Matches 11; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

QY 7 EGSSSQCCROEVQRKDLSSCERYL 32
 Db 33 EENQGEGRGRRERPEKRLDECRCREQAERR 59

RESULT 7
 Q9SQH1

ID Q9SQH1 PRELIMINARY; PRT; 160 AA.
 AC Q9SQH1;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ALLERGEN.
 GN ARA H 7.
 OS Arachis hypogaea (Peanut).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
 OC Arachis.
 OX NCBI_TaxID=3818;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VIRGINIA; TISSUE-SEED;
 RX MEDLINE=99406463; Pubmed=10474031;
 RA Kleber-Janke T.; Crameri R.; Appenzeller U.; Schlaak M.; Becker W.M.;
 RT "Selective cloning of peanut allergens, including profilin and 2S
 RT albumins, by phage display technology.";
 RL Int. Arch. Allergy Immunol. 119:265-274(1999).
 DR EMBL; AF091737; AAD56719.1; -
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_aml1_inh.
 DR Pfam: PF00234; tryp_alpha_aml1; 1.
 DR SMART; SM00499; AAI; 1.
 DR SEQUENCE 160 AA; 18417 MW; 9F94CEB68080D4C CRC64;

Query Match 33.1%; Score 61.5; DB 10; Length 160;
 Best Local Similarity 35.3%; Pred. No. 0.53;
 Matches 12; Conservative 11; Mismatches 8; Indels 3; Gaps 1;

QY 2 SKGERGSSSQ--QCRQEVQRKDLSSCERYL 32
 Db 30 SRGSRMDAPSRGDDCCQRLGRANLRPCHEMR 63

RESULT 8
 Q9AU64 PRELIMINARY; PRT; 572 AA.
 AC Q9AU64;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE 7S GLOBULIN.
 GN GLO7A.
 OS Elaeis guineensis var. tenera (Oil palm).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Arecaeae; Arecoidae;
 OC Cocoeae; Elaeidinae; Elaeis.
 OX NCBI_TaxID=51953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Morcillo F.; Hartmann C.; Duval Y.; Tregear J.;
 RT "Regulation of 7S globulin gene expression in zygotic and somatic
 RT embryos of oil palm.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF250228; AAK28402.1; -
 DR HSSP; P02853; 2PHL.
 DR InterPro: IPR001113; Seedstore_7s.
 DR Pfam; PF00546; Seedstore_7s; 1.
 DR Pfam; PF02808; Seedstore_7s-C; 1.
 DR SEQUENCE 572 AA; 66364 MW; 606A071866A6FB3B CRC64;

Query Match 32.8%; Score 61; DB 10; Length 572;
 Best Local Similarity 36.4%; Pred. No. 2.3;
 Matches 12; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

QY 5 EREGSSQCCROEVQRKDLSSCERYLRSSRR 37
 Db 74 ERKGQGEGRGRRERPEKRLDECRCREQAERR 106

RESULT	9			
09JGS8				
ID	09JGS8	PRELIMINARY;	PRT;	765 AA.
AC	09JGS8;			
DT	01-OCT-2000 (Tremblrel, 15, Created)			
DT	01-OCT-2000 (Tremblrel, 15, last sequence update)			
DT	01-DEC-2001 (Tremblrel, 19, last annotation update)			
DE	PORF1.			
OS	TT virus.			
OC	Viruses; ssDNA viruses; unclassified ssDNA viruses.			
OX	NCBI_TaxID=68887;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20417334; Pubmed=10963344;			
RA	Tanaka Y., Orito E., Ohno T., Nakano T., Hayashi K., Kato T.,			
RA	Mukaike M., Iida S., Mizokami M.;			
RT	"Identification of a 23kDa protein encoded by putative open reading			
RT	frame 2 of TT virus (TTV) genotype I different from the other			
RT	genotypes.";			
RL	Arch. Virol. 145:1385-1398(2000).			
RL	EMBL: AB030487; BAA90406.1; -.			
DR	InterPro: IPR004219: TT_ORF1.			
DR	Pfam: PF02956; TT_ORF1; 1.			
SQ	SEQUENCE 765 AA; 90222 MW; E845A26B69D90707 CRC64;			

Query Match	32.8%	Score 61	DB 12	Length 765
Best Local Similarity	46.9%	Pred. No. 3		
Matches 15, Conservative	7	Mismatches 8	Indels 2	Gaps 1

```
QY 1 ESKGEREGSSSQCRQEVQRKDISCERYLRQ 32
    ||:| ||:| ||:| ||:| ||:| ||:|
Db 704 ESQKEHEGTLSSQIREQVQQKLG--RQURE 733
```

RESULT	10
P93198	
ID	P93198
AC	P93198;
DT	01-MAY-1997 (TREMblrel_03, Created)
DT	01-MAY-1997 (TREMblrel_03, Last sequence update)
DT	01-DEC-2001 (TREMblrel_19, Last annotation update)
DE	ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
OS	Juglans regia (English walnut).
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC	eurosid1: Fagales: Juglandaceae; Juglans.
OX	NCBI_TaxID=51240;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	SPRAIN-CV, SUNLAND;
RA	Teuber S.S., Dandekar A.M., Peterson W.R.:
RT	"Juglans regia 2S albumin seed storage protein precursor."
RL	Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
DR	EMBL; U068866; AAB41308.1; -.
DR	InterPro: IPR003612; AAI.
DR	InterPro: IPR001768; Cereal_tryp_aml1_inh.
DR	InterPro: IPR000480; Glutelin.
DR	Pfam: PF00234; tryp_alpha_aml1; 1.
DR	PRINTS: PR00211; GLUTELIN.
DR	SMART: SM00499; AAI; 1.
FT	NON_TER 1
SQ	SEQUENCE 139 AA; 16373 MW; 02D0E55E67164F23 CRC64;

Query Match	32.5%	Score 60.5	DB 10	Length 139
Best Local Similarity	50.0%	Pred. No. 0.64		
Matches 11; Conservative	7	Mismatches	3	Indels 1; Gaps 14

14 CR0EV0R-KDLSCERYLRQSS 34

Db 39 CREQIQRQNLNHCQYLLRQS 60

RESULT	ID	Q9EXV9	PRELIMINARY;	PRT;	471 AA.
11	Q9EXV9				
AC	Q9EXV9				
DC	Q9EXV9				
D6	01-MAR-2001	(TREMblrel, 16, Created)			
D7	01-MAR-2001	(TREMblrel, 16, Last sequence update)			
D7	01-DEC-2001	(TREMblrel, 19, Last annotation update)			
DE	GDP-MANNOSE	PYROPHOSPHORYLASE.			
GN	MANC.				
OS	Salmonella enterica subsp. enterica.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Salmonella.				
OX	NCBI_TaxID=59201;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=M252;				
RX	MEDLINE=M135136; PubMed=11238967;				
RA	Jensen S.O., Reeves P.R.;				
RT	"Molecular evolution of the GDP-mannose pathway genes (manB and manC)				
RT	in Salmonella enterica.";				
RL	Microbiology 147:599-610(2001).				
DR	EMBL: AY012201; AAC41707.1; -				
DR	InterPro: IPR001538; MannoseP_isomer.				
DR	InterPro: IPR001825; NTP_transferase.				
DR	Pfam: PF00483; MannoseP_isomer; 1.				
DR	Pfam: PF00483; NTP_transferase; 1.				
DR	ProDom: PD002664; MannoseP_isomer; 1.				
SO	SEQUENCE	471 AA; 52962 MW; E3CB9A2CAFE36A90 CRC64;			

Query Match Similarity	31.7%	Score 59	DB 2	Length 471
Best Local Similarity	38.7%	Pred. No.	3.5	
Matches 12	Conservative 7	Mismatches 12	Indels 0	Gaps 0

QY 3 KGEREGSSSQCCRQEVORKDLSSCERYLRQS 33
 157 KGERVNDSSFQVESFVEKEPKDIETAKDYLRQN, 187

RESULT	12	
09JGSS5		
ID	09JGSS5	PRELIMINARY;
AC	09JGSS5;	PRT; 765 AA.
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	PORF1.	
OS	TT virus.	
OC	viruses; ssDNA viruses; unclassified ssDNA viruses.	
OX	NCBI_TaxID=68867;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20417334; PubMed=10963344;	
RA	Tanaka Y., Orito E., Ohno T., Nakano T., Hayashi K., Kato T.,	
RT	Mukaida M., Iida S., Murokami M.;	
RT	"Identification of a 23kDa protein encoded by putative open reading	
RT	frame 2 of TT virus (TTV) genotype 1 different from the other	
RT	genotypes";	
RL	Arch. Virol. 145:1385-1398(2000).	
RL	EMBL; AB030448; BAA90409.1; -	
DR	InterPro: IPR004219; TT_ORF1.	
DR	Pfam: PF02956; TT_ORF1, 1.	
Q	SEQUENCE 765 AA; 90275 MW; 707029EBA5829B5F CRC64;	

Query Match	31.7%;	Score 59;	DB 12;	Length 765;
Best Local Similarity	43.8%;	Pred. No. 5.7;		
Matches 14;	Conservative	9;	Mismatches 7;	Indels 2;
				Gaps 1;

QY 1 ESKGEREGSSQCCRQEVQRKDLSSCERYLRQ 32

Db 704 ESQKQEGTISQDLREQLQKLG--RQURE 733

RESULT 13
0941J9 PRELIMINARY; PRT; 462 AA.

AC 0941J9; 01-DEC-2001 (TREMUREL. 19, Created)
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Triticum.
OX NCBI_Taxid=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. S-149;
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. S-149;
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314772; CAC40684.1; -.
KW Signal.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SIGNAL 22 462
SQ SEQUENCE 462 AA; 49165 MW; 79C3BD615F91BF12 CRC64;

Query Match 31.2%; Score 58; DB 10; Length 462;
Best Local Similarity 35.7%; Pred. No. 4.8;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 5 EREGSSQOCROFYQRDLSSCERYLRQ 32

Db 22 EGEASGQLCCERLEACRQIVDQ 49

RESULT 14
0941I0 PRELIMINARY; PRT; 743 AA.
AC 0941I0; 01-DEC-2001 (TREMUREL. 19, Created)
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Secale.
OX NCBI_Taxid=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314776; CAC40677.1; -.
KW Signal.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SIGNAL 22 743
SQ SEQUENCE 743 AA; 78890 MW; A065980F7F2B0205 CRC64;

Query Match 31.2%; Score 58; DB 10; Length 743;
Best Local Similarity 35.7%; Pred. No. 7.7;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 5 EREGSSQOCROFYQRDLSSCERYLRQ 32

Db 22 EGEASGQLCCERLEACRQIVDQ 49

RESULT 15
0941I2 PRELIMINARY; PRT; 754 AA.
AC 0941I2; 01-DEC-2001 (TREMUREL. 19, Created)
DT 01-DEC-2001 (TREMUREL. 19, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HIGH MOLECULAR WEIGHT GLUTENIN SUBUNIT X PRECURSOR.
GN GLU-1R.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Trilicaceae; Secale.
OX NCBI_Taxid=4550;
RN [1]
RP SEQUENCE FROM N.A.
RA de Bustos A., Jouve N.;
RT "Characterization and analysis of new subunits-like wheat HMW-glutenin
RT encoded by Glu-1R gene of Secale cereale.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA de Bustos A., Rubio P., Jouve N.;
RT "Characterization of two gene subunits on the 1R chromosome of rye as
RT orthologs of each of the Glu-1 genes of hexaploid wheat.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314773; CAC40674.1; -.
KW Signal.
FT CHAIN 1 21 HMW GLUTENIN SUBUNIT X.
FT SIGNAL 22 754
SQ SEQUENCE 754 AA; 80179 MW; 89BAD795A3DC91BB CRC64;

Query Match 31.2%; Score 58; DB 10; Length 754;
Best Local Similarity 35.7%; Pred. No. 7.8;
Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

OY 5 EREGSSQOCROFYQRDLSSCERYLRQ 32

Db 22 EGEASGQLCCERLEACRQIVDQ 49

Search completed: September 9, 2002, 12:45:36
Job time: 293 sec

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 26-45;65-84 <HAR2>
C:Keywords: seed; storage protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-35/Domain: propeptide #status experimental <PRO>
F:36-141/Product: 2S albumin #status experimental <MAT>

Query Match 34.2%; Score 122; DB 2; Length 141;
Best Local Similarity 32.8%; Pred. No. 1.3e-05;
Matches 21; Conservative 18; Mismatches 23; Indels 2; Gaps 1;

QY 2 QOESQOLQCCNQVKQVRDECOCCEAIKTYAEDQIQGQLHGESESRVAQRAGEIVSSCGV 61
DB 77 RREGSEDECCEKLNDEECRCMDLEIARE--EORARQOEGKQMLQKARNLPSCMGI 134
QY 62 RCMR 65
DB 135 RPQR 138

RESULT 6
2S albumin - Brazil nut
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 13-Jan-1995 #sequence,revision 13-Jan-1995 #text,change 21-Jul-2000
C:Accession: S14947
R:Gander, E.S.; Holmstroem, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A:Reference number: S14946; MUID:91370890
A:Accession: S14947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <GAN>
A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA8363.1; PID:g17713
C:Genetics:
A:Introns: 67/3
C:Superfamily: wheat alpha-amylase inhibitor

Query Match 31.8%; Score 113.5; DB 2; Length 154;
Best Local Similarity 28.6%; Pred. No. 0.00011;
Matches 18; Conservative 21; Mismatches 21; Indels 3; Gaps 1;

QY 4 ESQOLQCCNQVKQVRDECOCCEAIKTYAEDQIQGQLHGESESRVAQRAGEIVSSCGV-- 61
DB 83 EEPHLDCECEQLERMDENCRCEGLRMMLRRQREMELEGEMQRLMKRAENLISRCNLS 142
QY 62 -RC 63
DB 143 QRC 145

RESULT 7
NMKUI
2S albumin 1 precursor - Arabidopsis thaliana
N:Alternate names: seed storage protein AT251
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 31-Mar-1992 #sequence,revision 31-Mar-1992 #text,change 03-Dec-1999
R:Kreibers, E.; Herdies, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; Se
Plant Physiol. 87, 859-866, 1988
A:Title: Determination of the processing sites of an Arabidopsis 2S albumin and characte
A:Reference number: JA0161
A:Accession: JA0161
A:Molecule type: DNA
A:Residues: 1-164 <KRE>
A:Cross-references: GB:M22032; NID:g166609; PIDN:AAA32743.1; PID:g166614
A:Accession: PS0282
A:Molecule type: protein
A:Residues: 38-73;84-162 <KR2>

R:Conceicao, A.D.S.; Kreibers, E.
submitted to the EMBL Data Library, July 1993
A:Description: Tentative title: a coryledon regulatory region is responsible for the

A:Reference number: S34674
A:Accession: S34676
A:Molecule type: DNA
A:Residues: 1-164 <CON>
A:Cross-references: EMBL:Z24745; NID:g395203; PIDN:CAA80870.1; PID:g395204
R:Bevan, M.; Van Der Schuren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.;
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15484
A:Accession: T06044

A:Molecule type: DNA
A:Residues: 1-164 <BEV>
A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.90
A:Experimental source: cultivar Columbia; BAC clone T24A18
C:Genetics:
A:Gene: T24A18.90
A:Map position: 4
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-164/Product: 2S albumin 1 proprotein #status predicted <AT2>
F:38-73/Product: 2S albumin 1 small chain #status experimental <SMC>
F:84-162/Product: 2S albumin 1 large chain #status experimental <LAC>

Query Match 30.8%; Score 110; DB 1; Length 164;
Best Local Similarity 34.4%; Pred. No. 0.00026;
Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

QY 1 QOESQOLQCCNQVKQVRDECOCCEAIKTYAEDQIQGQLHGESESRVAQRAGEIVSSCG 60
DB 87 QOEDQOLFQCCNCELROBEDPCVPTLKQAKAVRLQGHOPMQRKTYQAKHLPNVDC 146
QY 61 V 61
DB 147 I 147

RESULT 8
T09252
seed storage protein EMB25 - white spruce
C:Species: Picea glauca (white spruce)
C:Date: 11-Jun-1999 #sequence,revision 11-Jun-1999 #text,change 11-Jun-1999
C:Accession: T09252
R:Dong, J.Z.; Dunstan, D.I.
submitted to the EMBL data library, June 1996
A:Description: Gene expression during somatic embryogenesis.
A:Reference number: Z16588
A:Accession: T09252
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-165 <DON>
A:Cross-references: EMBL:L47745; NID:g1350513; PID:g1350514
C:Genetics:
A:Gene: EMB25

Query Match 30.1%; Score 107.5; DB 2; Length 165;
Best Local Similarity 26.6%; Pred. No. 0.00047;
Matches 21; Conservative 20; Mismatches 21; Indels 17; Gaps 1;

QY 1 QOESQOLQCCNQVKQVRDECOCCEAIKTYAED-----QIQGQLHGE 43
DB 66 EEPDQPSERCCEELQKSPQCRQATQRTLEDVFMDSQDGAPLNRQRGRGQGRGM 125
QY 44 ESERYAQRAGEIVSSCGVR 62
DB 126 EEEVVRARAEELPMTCNVR 144

RESULT 9


```

R.Kriebbers', E., Herdies, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; S
Plant Physiol. 87, 859-866, 1988
A:Title: Determination of the processing sites of an Arabidopsis 2S albumin and characte
A:Reference number: JA0161
A:Accession: JA0163
A:Molecule type: DNA
A:Residues: 1-164 <KRE>
A:Cross-references: GB:M2035; NID:g166611; PIDN:AAA32745.1; PID:g166616
R.Conceliao, A.D.S.; Kriebbers, E.
submitted to the EMBL Data Library, July 1993
A:Description: Tentative title: a cotyledon regulatory region is responsible for the diffe
A:Reference number: S34674
A:Accession: S34674
A:Molecule type: DNA
A:Residues: 1-164 <CON>
A:Cross-references: EMBL:T24744; NID:g395200; PIDN:CAA80868.1; PID:g395201
R.Itsugita, A.; Kano, M.; Kawakami, M.; Ohki, Y.
submitted to JIPID, December 1995
A:Description: Two dimensional electrophoresis of plant proteins and standardization of
A:Reference number: PN0173
A:Accession: PN0174
A:Molecule type: protein
A:Residues: 81-94 <TSU>
A:Experimental source: seeds
R.Beyran, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z13464
A:Accession: T06046
A:Molecule type: DNA
A:Residues: 1-164 <BEV>
A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP-T24A18.110
A:Experimental source: cultivar Columbia; BAC clone T24A18
C.Genetics:
A:Gene: T24A18.110
A:Map position: 4
A:Superfamily: wheat alpha-amylase inhibitor
C.Keywords: seed; storage protein
F.1-21/Domain: signal sequence #status predicted <SIG>
F.22-16/Product: 2S albumin 3 proprotein #status predicted <AT2>
F.38-72/Product: 2S albumin 3 small chain #status predicted <SMC>
F.82-162/Product: 2S albumin 3 large chain #status predicted <LMC>

Query Match      28.3%, Score 101, DB 1, Length 164;
Best Local Similarity 36.6%, Pred. No. 0.0022;
Matches 26; Conservative 11; Mismatches 28; Indels 6; Gaps 3;

Oy   2 OOESEOLOCCNQGVQVDECOCEAIKYIAEDQIQGGDLHE-ESERYAQAGETIVSSCG 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   86 QOGYQLDQQCCNEHQEPVCVPTLKQAARAVSLQG-HGPFSRKITYQSAYLPNTCK 144

Oy   61 VR---CMROT 67
    :: | | |
Db   145 IQQVGECPEPQT 155

RESULT 14
B23617
conglutin delta-2 large chain - narrow-leaved blue lupine
C:Species: Lupinus angustifolius (narrow-leaved blue lupine)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 11-Jan-2000
C:Accession: B23617
R.Lilley, G.G.; Inglis, A.S.
FEBS Lett. 195, 235-241, 1986
A:Title: Amino acid sequence of conglutin delta, a sulfur-rich seed protein of Lupinus e
A:Reference number: A91358
A:Accession: B23617
A:Molecule type: protein
A:Residues: 1-80 <LIU>
C:Superfamily: soybean 2S albumin

27.5%, Score 98; DB 2; Length 80;

```

[illegible]

Search completed: September 9, 2002, 12:43:30
Job time: 222 sec

```

Query Match      26.8%; Score 95.5; DB 2; Length 162;
Best Local Similarity 25.6%; Pred. No. 0.0081;
Matches    21; Conservative   21; Mismatches   21; Indels   19; Gaps   4;

QY       1  QOQESQOLQCCNQKQVRDECQCEAIRYI-----AEDQIQ--GQLHG---ESEE 46
          ::: | : ||::: : :| | | : :|| : | | : | :
Db        66  ERRRQPSESCCNELERMSPOCRCPAQIDVLDSASFMDSEDALNQRGRREGRRREQ 125
          E R R R Q P S E S C C N E L E R M S P O C R C P A Q I D V L D S A S F M D S E D A L N Q R G R R E G R R E Q

QY       47  RVAQRAGETVSSCGY-----RC 63
          ::||| : :| ||
Db       126  EMAERRAYLPDTCNVQESPRLC 147
          : : | | | : : | | |

Search completed: September 9, 2002, 12:43:30
Job time: 222 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:39:48 ; Search time 45.25 Seconds
(without alignments)
78.570 Million cell updates/sec

Title: US-09-913-351-2

Perfect score: 186

Sequence: 1 ESKGERESSSQGCRGEVQRKDLSSCERYLRQSSSR 37

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	100.0	258	1 RZCS	2S seed storage pr
2	73.5	40.6	141	2 T10257	2S albumin precurs
3	74	39.8	323	2 S38887	2S albumin - commo
4	70.5	37.9	154	2 S14947	2S albumin - Brazil
5	63.5	34.1	28	2 A25802	2S seed storage pr
6	62	33.3	749	2 S50095	splicing regulator
7	62	33.3	798	2 T15336	hypothetical prote
8	61.3	33.1	146	2 S14946	2S seed storage pr
9	59	31.7	295	2 S01062	2S seed storage pr
10	58.5	31.5	32	2 B59346	seed storage prote
11	58	31.2	148	2 S46514	putridolone-b pre
12	57	30.6	153	2 A33090	conjugulin delta pr
13	57	30.6	699	2 E84565	hypothetical prote
14	56	30.1	37	2 A23617	conjugulin delta-2
15	55	29.6	241	2 T27917	hypothetical prote
16	55	29.6	648	2 S04832	glutennin high mole
17	55	29.6	660	2 A24266	glutennin high mole
18	55	29.6	705	2 S18733	glutennin high mole
19	55	29.6	4815	2 JN0689	glutennin, high mole
20	54	29.0	417	2 T47616	hypothetical prote
21	54	29.0	1377	2 T51447	transcription regu
22	53.5	28.8	100	2 G88637	protein F53H1.4 [i
23	53.5	28.8	104	2 S48180	mabninin IV - Yun
24	53.5	28.8	104	2 S48178	mabninin III - Yun
25	53.5	28.8	838	1 E872HW	glutennin, high mole
26	53.5	28.8	848	1 S02262	glutennin high mole
27	53	28.5	215	1 A47118	periplasmic ribonu
28	53	28.5	301	2 E71558	probable fad synth
29	53	28.5	374	2 T33328	hypothetical prote

30	53	28.5	518	2 D84977	glutamate--cystein
31	53	28.5	685	2 B82606	conjugal transfer
32	53	28.5	815	2 B30843	glutennin high mole
33	53	28.5	830	2 S15720	glutennin high mole
34	53	28.5	3844	2 T18402	asparagine/asparta
35	52	28.0	722	2 T02421	hypothetical prote
36	52	28.0	2061	2 T13751	transcription fact
37	51.5	27.7	104	2 S48176	mabninin I-1 - Yun
38	51.5	27.7	161	2 UC4966	high-molecular-wet
39	51.5	27.7	165	2 T09252	seed storage prote
40	51.5	27.7	418	2 H83126	probable secretion
41	51.5	27.7	655	2 T28885	hypothetical prote
42	51	27.4	60	2 S29179	high-molecular-wet
43	51	27.4	63	2 S29177	high-molecular-wet
44	51	27.4	187	2 B83467	hypothetical prote
45	51	27.4	302	2 T13457	hypothetical prote

ALIGNMENTS

RESULT 1
RZCS
2S seed storage protein precursor - castor bean
N:Alternate names: 2S albumin precursor
C:Species: Ricinus communis (castor bean)
C:Date: 14-Nov-1983 #sequence, revision 08-Feb-1996 #text-change 18-Jun-1999
C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222
R:Irwin, S.D.; Keen, J.N.; Flindlay, J.B.C.; Lord, J.M.
Mol. Gen. Genet. 222, 400-408, 1990
A:Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro
A:Reference number: S11500; MUID:91109729
A:Accession: S11500
A:Molecule type: mRNA
A:Residues: 1-13, 'E', '15-73, 'T', '75-258 <IR2>
A:Experimental source: clone 14g4
A:Accession: S11501
A:Molecule type: mRNA
A:Residues: 'M', '4', 'LS', '7-13, 'F', '15-21 <IR>
A:Experimental source: clone 10a12
A:Accession: S27221
A:Molecule type: protein
A:Residues: 'X', '37, 'X', '39-45, 'X', '158-161, 'X', '163-174, 'X' <IRA>
R:Shacter, F.S.; Li, S.S.U.
J. Biol. Chem. 257, 14753-14759, 1982
A:Title: Amino acid sequence of small and large subunits of seed storage protein from
A:Reference number: A92357; MUID:83082772
A:Accession: A01328
A:Molecule type: protein
A:Residues: 157-190; 194-221, 'Q', '223-225, 230-233, 'N', '235-254, 'Q', '256-258 <SRA>
A:Note: 230-Ser was also found
A:Note: there is considerable similarity between residues 181-231 of this protein and
R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.
Biochem. J. 213, 543-545, 1983
A:Title: Structural relationships between barley (Hordeum vulgare) trypsin inhibitor a
A:Reference number: A90322; MUID:83308577
A:Contents: annotation
A:Note: this protein is homologous with trypsin inhibitor from barley
C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and
C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-35/Domain: propeptide #status predicted <PRO>
 F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC
 F:87-156/Product: probable 2S seed storage protein large 2 #status predicted <LCH
 F:157-190/Product: 2S seed storage protein small chain #status experimental <SML>
 F:194-255/Product: 2S seed storage protein large chain #status experimental <LNG>
 F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 186; DB 1; Length 258;
 Best Local Similarity 100.0%; Pred. No. 2,6e-16;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQSSRR 37
 DB 36 ESKGEREGSSQOCROEVQRKDLSCERYLRQSSRR 72

RESULT 2
 T10257
 2S albumin precursor - cucurbit
 N:Alternate names: prepro2S albumin
 C:Species: Cucurbita sp. (cucurbit)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 R:Hara-Nishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M.
 A:Accession: T10257; S19323
 A:Title: Vesicle transport and processing of the precursor to 2S albumin in pumpkin.
 A:Reference number: Z17000; MUID:94100993
 A:Accession: T10257
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA; protein
 A:Residues: 1-141 <HAR1>
 A:Cross-references: EMBL:D16560; NID:q459404; PIDN:BA03993.1; PID:q459405
 A:Experimental source: seed; storage deposition stage; cotyledon
 A>Note: soluble seed protein
 R:Hara-Nishimura, I.; Inoue, K.; Nishimura, M.
 F:BS Lett. 294, 89-93, 1991
 A:Title: A unique vacuolar processing enzyme responsible for conversion of several propt
 A:Reference number: S19323; MUID:92077151
 A:Accession: S19323
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 26-45;65-84 <HAR2>
 C:Keywords: seed; storage protein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-35/Domain: propeptide #status experimental <PRO>
 F:36-141/Product: 2S albumin #status experimental <MAT>

Query Match 40.6%; Score 75.5; DB 2; Length 141;
 Best Local Similarity 47.1%; Pred. No. 0.017;
 Matches 16; Conservative 7; Mismatches 10; Indels 1; Gaps 1;

OY 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQSS 34
 DB 31 ESKGEREGSSQOCROEVQRKDLSCERYLRQSS 63

RESULT 3
 S38887
 2S albumin - common sunflower (fragment)
 C:Species: Helianthus annuus (common sunflower)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-May-2000
 C:Accession: S38887
 R:Thyges, P.J.E.; Millichip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, F.
 submitted to the EMBL Data Library, November 1993
 A:Reference number: S38887
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-523 <THO>

A:Cross-references: EMBL:X76101; NID:q429181; PID:q429182
 C:Superfamily: gliadin

Query Match 39.8%; Score 74; DB 2; Length 323;
 Best Local Similarity 48.5%; Pred. No. 0.056;
 Matches 16; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

OY 6 REGSSQOCROEVQRKDLSCERYLRQ--SSSR 36
 DB 171 RSSSESQOCROGEIQRPVSCQRYVEQQLRSSR 203

RESULT 4
 S14947
 2S albumin - Brazil nut
 C:Species: Bertholletia excelsa (Brazil nut)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 R:Gander, E.S.; Holmstroem, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gr
 Plant Mol. Biol. 16, 437-448, 1991
 A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin
 A:Reference number: S14946; MUID:91370890
 A:Accession: S14947
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <GAN>
 A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA38363.1; PID:g17713
 C:Genetics:
 A:Introns: 67/3
 C:Superfamily: wheat alpha-amylase inhibitor

Query Match 37.9%; Score 70.5; DB 2; Length 154;
 Best Local Similarity 42.4%; Pred. No. 0.079;
 Matches 14; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

OY 1 ESKGEREGSSQOCROEVQRKDLSCERYLRQ 32
 DB 34 EOENPRGRSEOCROEMRQQLNHCWYLRQ 66

RESULT 5
 A25802
 2S seed storage protein small chain - Brazil nut
 C:Species: Bertholletia excelsa (Brazil nut)
 C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 19-May-2000
 C:Accession: A25802
 R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampalo, M.J.A.M.; Van Montagu, M.; Van
 Eur. J. Biochem. 159, 597-604, 1986
 A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil
 A:Reference number: A91173; MUID:87004679
 A:Accession: A25802
 A:Molecule type: protein
 A:Residues: 1-28 <AMP>
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: pyroglutamic acid
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 34.1%; Score 63.5; DB 2; Length 28;
 Best Local Similarity 54.5%; Pred. No. 0.13;
 Matches 12; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

OY 12 OOCROEVQRKDLSCERYLRQ 32
 DB 2 OOCROEVQRKDLSCERYLRQ 23

RESULT 6
 S50095
 Splicing regulator Ceswap - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C>Date: 01-Aug-1995 #sequence_revision 24-Jul-1997 #text_change 17-Mar-1999
C:Accession: S53765; S50095
R:Spikes, D.A.; Kramer, J.; Bingham, P.M.; van Doren, K.
Nucleic Acids Res. 22, 4510-4519, 1994
A:Title: SMAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a
A:Reference number: S50095; MUID:95061415
A:Accession: S53765
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-749 <SPI>
A:Cross-references: EMBL:U06932; NID:g459659; PID:g459660
A:Accession: S50095
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-551 <SP2>
A:Cross-references: EMBL:U06933
C:Genetics:
A:Introns: 56/2; 104/1; 434/3; 513/3
A>Note: this list of introns may be incomplete

Query Match 33.3%; Score 62; DB 2; Length 749;
Best Local Similarity 46.2%; Pred. No. 4.1;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EREGSSOOCROEVORKDLSCEERYL 30
DB 71 KRNGSPSEOCPTAEEDWCCEERYL 96

RESULT 7
T15336
Hypothetical protein B0336.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15336
R:Taich, A.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid B0336.
A:Reference number: Z18331
A:Accession: T15336
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-798 <TAI>
A:Cross-references: EMBL:U03205; NID:g912752; PID:g912761; PIDN:AA046835.1; CESP:B0336.9
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:B0336.9
A:Introns: 17/3; 60/2; 80/2; 128/1; 458/3; 537/3

Query Match 33.3%; Score 62; DB 2; Length 798;
Best Local Similarity 46.2%; Pred. No. 4.4;
Matches 12; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

OY 5 EREGSSOOCROEVORKDLSCEERYL 30
DB 95 KRNGSPSEOCPTAEEDWCCEERYL 120

RESULT 8
S14946
2S seed storage protein large chain - Brazil nut
N:Alternate names: albumin 2S precursor
C:Species: Bertholletia excelsa (Brazil nut)
C>Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 21-Jul-2000
C:Accession: S14946; S14479; S06252; S21640; B25802
R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
A:Reference number: S14946; MUID:91370890
A:Accession: S14946
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-146 <GAN>
A:Cross-references: EMBL:X54490; NID:g17710; PIDN:CAA38362.1; PID:g17711
A>Note: the authors translated the codon CTT for residue 13 as Val and GTC for residu
R:Bassuener, R.; Schlesier, B.
submitted to the EMBL Data Library, December 1990
A:Reference number: S14479
A:Accession: S14479
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BAS>
A:Cross-references: EMBL:X57027; NID:g17714; PID:g17715
R:Altenbach, S.B.; Pearson, K.W.; Leung, F.W.; Sun, S.S.M.
Plant Mol. Biol. 8, 239-250, 1987
A:Title: Cloning and sequence analysis of a cDNA encoding a Brazil nut protein except
A:Reference number: S06252
A:Accession: S06252
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-146 <ALT>
A:Experimental source: clone pHS-3
A>Note: part of this sequence, including the amino end of the large chain, was confir
R:Bassuener, R.; Schlesier, B.
submitted to the EMBL Data Library, December 1990
A:Reference number: S21640
A:Accession: S21640
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BA2>
A:Cross-references: EMBL:X57028; NID:g17716; PID:g17717
R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampato, M.J.A.M.; Van Montagu, M.; Van
Eur. J. Biochem. 159, 597-604, 1986
A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil
A:Reference number: A91173; MUID:87004679
A:Accession: B25802
A:Molecule type: protein
A:Residues: 70-90, 'E', 92-121, 'W', 123-125, 'L', 127-142 <AMP>
C:Genetics:
A:Introns: 60/3
C:Superfamily: wheat alpha-amylase inhibitor
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-46/Domain: propeptide #status predicted <PRO>
F:47-69/Product: seed storage protein small chain #status predicted <SCH>
F:70-146/Product: seed storage protein large chain #status experimental <LCH>

Query Match 33.1%; Score 61.5; DB 2; Length 146;
Best Local Similarity 40.7%; Pred. No. 1;
Matches 11; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

OY 7 EGGSSOOCROEVORKDLSCEERYLRQ 32
DB 33 EEENOECECEMOOROMLSHCRMYRQ 59

RESULT 9
S01062
2S seed storage protein precursor (clone Hag5) - common sunflower
N:Alternate names: 2S albumin storage protein
C:Species: Helianthus annuus (common sunflower)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 21-Jul-2000
C:Accession: S01062
R:Allen, R.D.; Cohen, E.A.; Vonder Haar, R.A.; Adams, C.A.; Ma, D.P.; Nessler, C.L.;
Mol. Gen. Genet. 210, 211-218, 1987
A:Title: Sequence and expression of a gene encoding an albumin storage protein in sun
A:Reference number: S01062; MUID:88142538
A:Accession: S01062
A:Molecule type: DNA
A:Residues: 1-295 <ALL>
A:Cross-references: EMBL:X06410; NID:g18799; PIDN:CAA29699.1; PID:g18800
A>Note: part of this sequence, including the amino end of the mature protein, was con
C:Genetics:
A:Introns: 192/2

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:42:39 : Search time 94.6 Seconds
(without alignments)
79.842 Million cell updates/sec

Title: US-09-913-351-3

Perfect score: 357
Sequence: 1 QQQESQQLQCCNQVKQVND.....AQRAGEIVSSCGVRCMKQTR 68

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_032802:*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	100.0	68	21	AAAB15422
2	121	33.9	65	21	AAAB15423
3	110	30.8	164	10	AAAP91892
4	110	30.8	164	10	AAAP96144
5	107.5	30.1	158	18	AAAW23588
6	106.5	29.8	158	18	AAAW23585
7	98	27.5	171	15	AAAR53579
8	98	27.5	184	21	AAI49926
9	96	26.9	184	12	AAAR14193
10	96	26.9	184	15	AAAR54981
11	96	26.9	184	16	AAAR76694

12	96	26.9	184	18	AAAM15415
13	96	26.9	184	22	AAAB30526
14	96	26.9	184	22	AAAB85931
15	96	26.9	189	9	AAAB81140
16	96	26.3	142	21	AAAR54980
17	94	26.3	142	21	AAAR84356
18	91.5	25.6	166	22	AAAB33600
19	91.5	25.6	166	22	AAAB05035
20	91.5	25.6	167	22	AAAB04710
21	91	25.5	184	19	AAAB30670
22	90.5	25.4	155	18	AAAW23418
23	90.5	25.4	158	18	AAAW23420
24	87.5	24.5	156	20	AAAY40973
25	87.5	24.5	157	18	AAAW24164
26	87.5	24.5	157	18	AAAW24153
27	87.5	24.5	157	20	AAAY15245
28	87.5	24.5	157	22	AAAB04707
29	87.5	24.5	166	20	AAAY40968
30	87.5	24.5	207	22	AAAB82383
31	86.5	24.2	155	18	AAAM10569
32	86.5	24.2	155	18	AAAW23587
33	85.5	23.9	93	12	AAAR14111
34	85.5	23.9	110	15	AAAR62762
35	85.5	23.9	111	19	AAAR62664
36	85.5	23.9	111	19	AAAR62669
37	85.5	23.9	140	10	AAAP91891
38	84.5	23.7	111	19	AAAR62721
39	84.5	23.7	111	19	AAAR62627
40	84.5	23.7	111	19	AAAR62628
41	84.5	23.7	111	19	AAAR62627
42	84.5	23.7	148	21	AAAY58049
43	84.5	23.7	149	19	AAAR58050
44	84.5	23.7	149	19	AAAR62722
45	84.5	23.7	149	19	AAAR62723

ALIGNMENTS

RESULT 1	
AAAB15422	standard; peptide; 68 AA.
AAAB15422;	
17-JAN-2001	(first entry)
Castor bean 2S albumin storage protein amino acids 87-154.	
Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;	
polysaccharide; mannose; galactose; castor bean; immune response; human;	
2S albumin storage protein; animal; tumour necrosis factor; mononuclear;	
phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.	
OS	Ricinus communis.
PN	WO200050087-A1.
PD	31-AUG-2000.
PF	21-OCT-1999; 99MO-ES00338.
PR	26-FEB-1999; 99ES-0000408.
PA	(INFA-) IND FARM CANTABRIA SA.
PI	Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;
PI	Piyel Rantieri JP, Gimenez Gallejo G, Matji Tuduri JA;
DR	WPI; 2000-558369/51.
PT	New glycoconjugate, useful for treating immunological disorders,
PT	comprises polysaccharide from Candida utilis and polypeptide from

PT Ricinus communis -
XX
XX
PS Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC ZACZBC(QERK)2(Z')(LIVM)ZCCC(Z')(QEH)(LV)Z6CZCZ2ZdGZCZ(VILM)ZICZg
CC where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 87-154 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from Candida utilis. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
XX
SQ Sequence 68 AA;

Query Match 100.0%; Score 357; DB 21; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.6e-36;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QOOSQOOLQCCNCKVQKVRDECCCAIKYIAEDQIOGOLGHESESRVAPRAGRIVSSCG 60
DB 1 qqqesqgqgqcnvkvqrdecccaikylaedqigqgqIngesevraqagelvsqg 60
QY 61 VRCKRQTR 68
DB 61 vrcmrqr 68

RESULT 2
AAB15423
ID AAB15423 standard; peptide; 65 AA.
XX
AC AAB15423;
XX
DT 17-JAN-2001 (first entry)
XX
DE Castor bean 2S albumin storage protein amino acids 194-258.
XX
XX Antitumour; antimicrobial; immunostimulatory; glycoconjugate; mannose;
KW polysaccharide; mannose; galactose; castor bean; immune response; human;
KW 2S albumin storage protein; animal; tumour necrosis factor; mononuclear;
KW phagocyte; granulocyte; macrophage; tumour; immunogen; Candida utilis.
XX
OS Ricinus communis.
XX
PN W0200050087-A1.
XX
PD 31-AUG-2000.
XX
PF 21-OCT-1999; 99WC-ES00338.
XX
PR 26-FEB-1999; 99ES-0000408.
XX
XX (INFA-) IND FARM CANTABRIA SA.
XX
PA Briteva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;
PI Pivel Ranieri JF, Gimenez Gallego G, Matji Tuduri JA;
XX

DR WPI; 2000-558369/51.
XX
XX
PT New glycoconjugate, useful for treating immunological disorders,
PT comprises polysaccharide from Candida utilis and polypeptide from
PT Ricinus communis -
XX
XX
PS Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC ZACZBC(QERK)2(Z')(LIVM)ZCCC(Z')(QEH)(LV)Z6CZCZ2ZdGZCZ(VILM)ZICZg
CC where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 194-258 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from Candida utilis. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
XX
SQ Sequence 65 AA;

Query Match 33.9%; Score 121; DB 21; Length 65;
Best Local Similarity 34.5%; Pred. No. 1.8e-07;
Matches 20; Conservative 15; Mismatches 23; Indels 0; Gaps 0;

QY 4 ESQOLQCCNCKVQKVRDECCCAIKYIAEDQIOGOLGHESESRVAPRAGRIVSSCGV 61
DB 1 qerslrgccdhkqmgsgrcqglrqaiegsgqgqigqvfafraanlpsmcgv 58

RESULT 3
AAP91892
ID AAP91892 standard; protein; 164 AA.
XX
AC AAP91892;
XX
DT 29-APR-1990 (first entry)
XX
DE Amino acid sequence of 1kb fragment of the Arabidopsis thaliana
XX 2S-albumin gene.
XX
XX 2S-albumin; Arabidopsis thaliana; storage protein gene;
KW heterologous polypeptide.
KW
XX
OS Arabidopsis thaliana.
XX
FH Key
FH Peptide 1..37 Location/Qualifiers
FT /note="signal peptide" 38..73
FT /note="mature small subunit" 74..83
FT /note="processing site" 84..157
FT /note="mature large subunit"
XX
PN W08903887-A.
XX
XX 05-MAY-1989.
XX

PF 20-OCT-1998 88W0-EP00944.
 XX
 PR 20-OCT-1997 87EP-0402348.
 XX
 PR (PILAN-) PLANT GENETIC SYST.
 XX
 PS Vandekerckhove JS, Krebbers E, Botterman J, Leemans J;
 DR MPI: 1989-150783/20
 DB N-PSDB: AAN91699.
 XX
 PT Recombinant DNA expression in plants
 XX using modified storage protein genes for expressing
 XX heterologous polypeptide(s) in the seeds
 PS Fig 4: 31pp: English.
 CC The entire 2S-albumin storage protein precursor including
 CC signal peptide. It is to be inserted into plants under the control of
 CC a seed-specific promoter and expressed at high levels only or mostly
 CC in the seed-forming stage and produced mostly in the seeds.
 XX
 SQ Sequence 164 AA:
 Query Match 30.8%; Score 110; DB 10; Length 164;
 Best Local Similarity 34.4%; Pred. No. 1.2e-05;
 Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
 OY 1 OOOESQQLQCCNOVKQVDECCCEAIKRTIADQIQGQLHGESESRVQRAGEIYSSCG 60
 DB 87 gqgeqqlfqccneirgqeedcvcptlkgaakavrlqgqhgmqrklyqtakhpnvcd 146
 OY 61 V 61
 DB 147 I 147
 RESULT 5
 ID AAW23588 standard; protein: 164 AA.
 AC AAW23588
 XX
 DE 20-JAN-1994 (first entry)
 XX
 DB Sequence encoded by 1kb fragment containing the Arabidopsis thaliana 2S
 DB albumin gene and related elements.
 XX
 KW Transgenic plant; nutrition enrichment; storage protein.
 OS Arabidopsis thaliana
 FT Key Location/Qualifiers
 FT Region 38..83
 FT /Label=small subunit
 FT Region 84..164
 FT /Label=large subunit
 FT Peptide 1..73
 FT Protein 74..164
 XX
 PN AAW23588
 DB 20-OCT-1998
 XX
 PD 20-OCT-1998 88EP-0402650.
 XX
 PR 20-OCT-1998 88EP-0402611.
 XX
 PR (PILAN-) PLANT GENETIC SYST.
 XX
 PS Vandekerckhove JS, Krebbers E, Vandekerckhove JS;
 DR MPI: 1989-150783/22.
 DB N-PSDB: AAN90116.
 XX
 PT Transgenic plants contg. protein of high nutritional value -
 PT obtd. by inserting gene for 2S albumin modified for enrichment
 PT of specific amino acid(s)
 XX
 PS Example: Fig 4: 31pp: English.
 CC Prodn. of transgenic plants with increased nutritional value is claimed.
 CC The method comprises cultivating plants, obtd. over 1 or more generations
 CC from regenerated plant cells (or seeds) which include a nucleic acid
 CC encoding a modified form of a natural 2S albumin storage protein, under
 CC control of a promoter. The modified proteins are expressed at high
 CC levels, without any alteration to correct expression, processing and
 CC transport.
 XX
 SQ Sequence 164 AA:
 Query Match 30.8%; Score 110; DB 10; Length 164;
 Best Local Similarity 34.4%; Pred. No. 1.2e-05;
 Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
 OY 1 OOOESQQLQCCNOVKQVDECCCEAIKRTIADQIQGQLHGESESRVQRAGEIYSSCG 60
 DB 87 gqgeqqlfqccneirgqeedcvcptlkgaakavrlqgqhgmqrklyqtakhpnvcd 146
 OY 61 V 61
 DB 147 I 147
 RESULT 5
 ID AAW23588 standard; protein: 158 AA.
 AC AAW23588;
 XX
 DE 30-SEP-1997 (first entry)
 XX
 DE Mabinlin MBLIII from Capparis masakal.
 XX
 KW MBU, sweetener; transgenic plant; recombinant protein; food;
 KW beverage; animal feed; chewing gum; dental hygiene product;
 KW pharmaceutical.
 OS Capparis masakal.
 FT Key Location/Qualifiers
 FT Peptide 1..39
 FT /Label=Signal
 FT Protein 40..158
 FT /Label=MBLIII
 FT Region 40..70
 FT /Label=A_chain
 FT Region 71..86
 FT /note="14 amino acids evidently lost when mature
 FT dimer is formed"
 FT Region 87..158
 FT /Label=B_chain
 XX
 PN WO9700945-A1.
 DB 09-JAN-1997.
 XX
 PD 09-JAN-1997.
 XX
 PR 21-JUN-1996; 96WO-US10669.
 XX
 PR 23-JUN-1995; 95US-0000480.
 XX
 PA (UYHA-) UNIV HAWAII.
 PI Chen H, Hu Z, Sun SSM, Xiong L;


```

PR      23-OCT-1992;       92US-0965664.
XX
PA      (PION-) PIONEER HI-BRED INT INC.
XX
PI      Ballo B;
XX
DR      WPI: 1994-167470/20.
NR      N-PSDB: AAQ63710.
XX
PT      Improving the content of an amino acid in a seed storage protein
PT      to enhance nutritional value - by replacing non-conserved DNA
PT      with DNA encoding the amino acid
XX
PS      Disclosure; Fig 1; 33pp: English.
XX
CC      The sequence is that of a 2S seed storage protein with increased
CC      lysine content. The protein is synthesised from a set of partial
CC      DNA sequences capable of being assembled in a complementary over-
CC      lapping relationship to provide the complete DNA. The protein is
CC      translated as the straight chain shown and undergoes post translational
CC      modification to give the mature protein which has two subunits joined by
CC      a di-sulphide bond.
CX
SQ      Sequence 171 AA;

Query Match          27.5%; Score 98; DB 15; Length 171;
Best Local Similarity 30.3%; Pred. NO. 0.00037;
Matches 20; Conservative 16; Mismatches 24; Indels 6; Gaps 2;

QY      2 QOESQIQGCCNQVKNQVRDECCQE---AIKTYIADQIQGQLHGEES--ERVAGRAGEI 55
        |||||::||:::|| | | | | | | | | | | | | | | | | | | | | | | | |
Db      89 qqrplllqkccceqlkqmgsqcvcplkgaskavkgeeqgqggqgqgmvrkiyltkahl 148
QY      56 VSSCGV 61
        | : |
Db      149 pkrvcdl 154

RESULT      8
AAAY49926
ID      AAY49926 standard; Protein: 184 AA.
AC      AAY49926;
XX
DT      31-JAN-2000 (first entry)
XX
DE      Brassica campestris lambda-CGM1-2 napin protein sequence.
XX
KW      Brassica; regulation; development; identification; screening;
KW      light promoter; seed promoter; fruit promoter; regulatory region;
KW      napin; soybean; rapeseed; cotton; safflower; sunflower; expression;
KW      phenotype; modification.
XX
OS      Brassica campestris.
PN      US5981839-A.
PD      09-NOV-1999.
PE      07-MAR-1997;   97US-0812665.
XX
PR      17-JAN-1985;     85US-0692605.
PR      23-JAN-1988;     88US-0147781.
PR      29-APR-1988;     88US-0188361.
PR      02-NOV-1988;     88US-0267865.
PR      09-JUL-1990;     90US-0550804.
PR      10-AUG-1993;     93US-0105852.
PR      07-JUN-1995;     95US-0484941.
PR      31-JUL-1986;     86US-0891529.
PR      26-MAY-1987;     87US-0054369.
PR      28-JUL-1987;     87US-0078538.
PR      15-MAR-1987;     88US-0168190.
```

```

PR 21-MAY-1990; 90US-0526123.
PR 14-SEP-1990; 90US-0582241.
PR 08-AUG-1991; 91US-0742834.
XX
XX
PA (CALJ ) CALGENE LLC.
XX
PI Kridl JC, Knauf VC;
XX
XX WPI; 2000-012275/01.
DR N-PSDB; AA235691.
XX
XX New genetic constructs useful for transforming plants with a DNA
PT sequence of interest -
XX
XX Example 3; Fig 2; 65pp; English.
XX
XX The present invention describes genetic constructs (i) for transforming
CC plants with a DNA sequence of interest, comprising a promoter from a
CC napin, E49 or acyl carrier protein gene, a DNA sequence of interest and
CC flanking T-DNA. The constructs can be used for the expression of
CC heterologous genes in plants. They can be used for modifying the
CC phenotype of plants. The DNA sequence of interest may encode an enzyme
CC or may comprise an antisense sequence. The constructs can be used in
CC plants such as soybean, rapeseed, cotton, safflower or sunflower. The
CC present sequence represents the Brassica campestris lambda-CGN1-2 napin
CC protein.
CC
SQ Sequence 184 AA;

Query Match 27.5%; Score 98; DB 21; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0004;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

QY 1 000EESQ-----LQCCCNQYKQVRDCCCEAIK---YIAEDQI--QOG-QLHGEESE----- 46
   ||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 92 qqgqpgpqneippll|qgcncne|evcptLkgaakavxqrlrqqgqmqgqmqyvis 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 47 RYACRAGEIVSSCGVR 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 152 rltgtatthpractnrlr 167

RESULT 9
AAR14193
ID AAR14193 standard; Protein; 184 AA.
XX
XX AAR14193;
AC
XX
XX 11-DEC-1991 (first entry)
DT
XX
XX Napin 1-2 gene product.
DE
XX
XX Desaturase; expression cassette; B. campestris.
KM
XX
XX WO9113972-A.
PN
XX
XX 19-SEP-1991.
PD
XX
XX 14-MAR-1991; 91WO-U001746.
PF
XX
XX 14-NOV-1990; 90US-0615784.
PR 16-MAR-1990; 90US-0494106.
PR 13-AUG-1990; 90US-0567373.
XX
XX (CALG-) CALGENE INC.
PA
XX
XX Thompson G, Knauf V;
PI
XX
XX WPI; 1991-295627/40.
DR N-PSDB; AAQ13969.
XX
XX DNA encoding a plant desaturase - used for modifying the satd.

```



```

RESULT 12
AAW15415
ID AAW15415 standard; Protein; 184 AA.
XX
AC AAW15415;
XX
DT 05-JUN-1997 (first entry)
XX
DE Napin encoded by clone gamma-ACGN1-2.
XX
KW Plant seed cell; napin gene; acyl carrier protein gene; EA9 gene;
KM Brassica; fatty acid; rat thioesterase; acyl ACP;
XX phospholipid acyl desaturase.
XX
OS Brassica campestris.
XX
PN US5608152-A.
XX
PD 04-MAR-1997.
XX
PF 31-JUL-1986; 86US-0891529.
XX
PR 09-JUL-1990; 90US-0550804.
PR 31-JUL-1986; 86US-0891529.
PR 28-JUL-1987; 87US-0078538.
PR 25-JAN-1988; 88US-0147781.
PR 08-AUG-1991; 91US-0742834.
PR 30-MAY-1995; 95US-0453924.
XX
PA (CALJ ) CALGENE INC.
XX
PI Knauf VC, Kridl JC;
XX
DR WPI; 1997-164560/15.
DR N-PSDB; AAT65475.
XX
PT Transgenic Brassica plant and seeds - contain DNA construct for seed
PT specific expression of heterologous proteins
XX
PS Example 2; Column 23-28; 48pp; English.
XX
CC The sequences given in AAW15415-17 represent proteins encoded by DNA
CC fragments which were used in the construction of the DNA construct
CC of the invention. The DNA construct comprises, in the 5' to 3'
CC direction of transcription:
CC (a) a transcriptional initiation region from a gene that encodes a
CC product preferentially expressed in a plant seed cell as compared to
CC other plant cells;
CC (b) a DNA sequence of interest other than the native coding sequence
CC of the gene; and
CC (c) a transcriptional termination region;
CC where the gene is a napin gene, an acyl carrier protein gene or an EA9
CC gene. This construct may be used in the production of a Brassica plant
CC which may be modified such that heterologous proteins are produced in
CC the seeds, such as fatty acids, rat thioesterase, acyl ACP or
CC phospholipid acyl desaturase.
XX
SQ Sequence 184 AA:

Query Match 26.9%; Score 96; DB 18; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0007;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

```

```

RESULT 13
AAW30526
ID AAW30526 standard; Protein; 184 AA.
XX
AC AAW30526;
XX
DT 26-OCT-1998 (first entry)
XX
DE Brassica campestris napin seed storage protein.
XX
KW Transgenic plant; tissue-specific gene expression; napin gene;
KM promoter; seed storage protein.
XX
OS Brassica campestris.
XX
PN US5753475-A.
XX
PD 19-MAY-1998.
XX
PF 10-AUG-1993; 93US-0105852.
XX
PR 10-AUG-1993; 93US-0105852.
PR 17-JAN-1985; 85US-0692605.
PR 31-JUL-1986; 86US-0891529.
PR 26-MAY-1987; 87US-0054369.
PR 28-JUL-1987; 87US-0078538.
PR 25-JAN-1988; 88US-0147781.
PR 15-MAR-1988; 88US-0168190.
PR 29-APR-1988; 88US-0188361.
PR 02-NOV-1988; 88US-0267685.
PR 21-MAY-1990; 90US-0526123.
PR 14-SEP-1990; 90US-0582241.
PR 08-AUG-1991; 91US-0742834.
XX
PA (CALJ ) CALGENE INC.
XX
PI Houck CM;
XX
DR WPI; 1998-311403/27.
DR N-PSDB; AAV36968.
XX
PT Transformation of plants - with regulatory sequence containing
PT constructs for tissue specific expression of genes
XX
PS Example 3; Fig 2E-G; 67pp; English.
XX
CC This is the napin seed storage protein of Brassica campestris,
CC encoded by genomic DNA in clone lambda CGN1-2 (see AAV36968). The
CC promoter region of the napin gene can be used in a novel method for
CC producing a plant with a regulatable phenotype. In this method,
CC regulatory regions from plant genes expressed during a particular
CC developmental stage or in a specific tissue are identified by cDNA
CC screening. The regulatory regions (and T-DNA and Ti or Ri plasmids)
CC are manipulated for use with foreign sequences for introduction
CC into plant cells to provide transformed plants having a phenotypic
CC property that can be modulated. The invention is exemplified with
CC light, seed and fruit-specific promoters. Also claimed are methods
CC for altering the phenotype of fruit tissue as distinct to other
CC plant tissue, modifying the genotype of a plant to impart a desired
CC characteristic to fruit, modifying transcription in fruit tissue,
CC and expressing a heterologous DNA sequence of interest specifically
CC in fruit tissue. The method provides transformation without gall
CC formation in plants which have historically not been Agrobacterium
CC hosts.
XX
SQ Sequence 184 AA:

Query Match 26.9%; Score 96; DB 19; Length 184;
Best Local Similarity 35.5%; Pred. No. 0.0007;
Matches 27; Conservative 12; Mismatches 23; Indels 14; Gaps 5;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:44:12 : Search time 34.92 Seconds
(without alignments)
47.564 Million cell updates/sec

Title: US-09-913-351-3

Perfect score: 357
Sequence: 1 QOQESQQLQCCNQVKQVARD.....AQRAGEIVSSCGVRCMRQRR 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/PCFUS.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107.5	30.1	158	2	US-08-670-186-6
2	106.5	29.8	158	2	US-08-670-186-4
3	96	26.9	184	1	US-08-453-924-3
4	90.5	25.4	155	2	US-08-618-911-2
5	90.5	25.4	158	2	US-08-618-911-6
6	86.5	24.2	155	2	US-08-670-186-2
7	83.5	23.4	158	2	US-08-618-911-4
8	83.5	23.4	158	2	US-08-938-675A-2
9	74	20.7	369	2	US-08-991-300-2
10	73.5	20.6	373	1	US-08-625-309A-2
11	71	19.9	587	1	US-07-955-805A-23
12	66	18.5	360	2	US-08-531-927B-2
13	66	18.5	360	4	US-09-041-886-13
14	65	18.2	1642	1	US-08-447-411-45
15	65	18.2	1642	2	US-08-662-227-2
16	65	18.2	1642	2	US-09-017-947-2
17	65	18.2	1648	2	US-08-662-227-35
18	65	18.2	1648	4	US-09-017-947-35
19	63.5	17.8	419	2	US-08-686-417-3
20	63.5	17.8	428	1	US-08-190-802A-29
21	63.5	17.8	428	4	US-08-477-346-29
22	63.5	17.8	428	4	US-08-473-089-29
23	62	17.4	124	1	US-08-216-846-2
24	62	17.4	124	1	US-08-782-177-3
25	62	17.4	261	1	US-08-245-688-2
26	62	17.4	261	1	US-08-245-688-4
27	62	17.4	261	1	US-08-245-688-6

28	62	17.4	261	1	US-08-245-688-8	Sequence 8, Appli
29	62	17.4	261	1	US-08-245-688-10	Sequence 10, Appli
30	62	17.4	261	1	US-08-245-688-12	Sequence 12, Appli
31	62	17.4	2101	1	US-08-466-390-4	Sequence 4, Appli
32	62	17.4	2101	1	US-08-470-950-4	Sequence 4, Appli
33	62	17.4	2101	1	US-08-467-781-4	Sequence 4, Appli
34	62	17.4	2101	1	US-08-195-487-4	Sequence 4, Appli
35	62	17.4	2101	2	US-08-483-924-4	Sequence 4, Appli
36	62	17.4	2101	4	US-09-452-284-1	Sequence 1, Appli
37	62	17.4	2101	5	PCT-US93-06160-4	Sequence 1, Appli
38	61.5	17.2	1651	1	US-08-447-411-2	Sequence 2, Appli
39	61	17.1	1099	4	US-09-442-100-2	Sequence 2, Appli
40	59.5	16.7	122	1	US-08-782-177-2	Sequence 2, Appli
41	59.5	16.7	124	1	US-08-216-846-1	Sequence 1, Appli
42	59.5	16.7	124	1	US-08-782-177-1	Sequence 1, Appli
43	59	16.5	700	2	US-08-568-459A-10	Sequence 10, Appli
44	59	16.5	700	2	US-08-487-826B-10	Sequence 10, Appli
45	59	16.5	2182	2	US-08-487-826B-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-08-670-186-6
Sequence 6, Application US/08670186
Patent No. 5839343
GENERAL INFORMATION:
APPLICANT: SUN, SAMUEL S. M.
APPLICANT: XIONG, LIWEN
APPLICANT: HU, ZHONG
APPLICANT: CHEN, HANG
TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESS: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,186
FILING DATE: 21-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 23461-20007.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-670-186-6

Query Match 30.1% Score 107.5; DB 2; Length 158;
Best Local Similarity 32.8%; Pred. No. 2.7e-06;
Matches 22; Conservative 13; Mismatches 27; Indels 5; Gaps 2;
QY 2 QOQESQQLQCCNQVKQVARDCCCEAIRYIARDQIQGOLHG-EESERYAQRAGEIVSSCG 60
11 :||||:| | : : | :||:| : :| :| :|


```

; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; City: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-618-911-2

```

```

Query Match          25.4%; Score 90.5; DB 2; Length 155;
Best Local Similarity 25.0%; Pred. No. 0.00036;
Matches 16; Conservative 25; Mismatches 16; Indels 7; Gaps 3;

```

```

QY 1 QQESQQLQCCNQVQKVRDE-CQCEAIKYIAEDQIQGQLHGESESVAVRAGEIVSSC 59
Db 84 EEERGHMKCKSEMSLKSPKCKCKALQKIMQNSE--QLEGEKKQMERH---LMNL 137
QY 60 GYRC 63
Db 138 AIRC 141

```

```

RESULT 5
US-08-618-911-6
; Sequence 6, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig
; APPLICANT: Coughlan, Sean
; APPLICANT: Hu, David
; TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
; SEEDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; City: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444

```

```

; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-618-911-6

```

```

Query Match          25.4%; Score 90.5; DB 2; Length 158;
Best Local Similarity 25.4%; Pred. No. 0.00037;
Matches 15; Conservative 24; Mismatches 17; Indels 3; Gaps 2;

```

```

QY 2 QQESQQLQCCNQVQKVRDE-CQCEAIKYIAEDQIQGQLHGESESVAVRAGEIVSSC 59
Db 88 EEERGHMKCKSEMSLKSPKCKCKALQKIMQNSE--ELEEKMKMKEMLMNLATMC 144

```

```

RESULT 6
US-08-670-186-2
; Sequence 2, Application US/08670186
; Patent No. 5859343
; GENERAL INFORMATION:
; APPLICANT: SUN, SAMUEL S.M.
; APPLICANT: XIONG, LIWEN
; APPLICANT: HU, ZHONG
; APPLICANT: CHEN, HANG
; TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500
; City: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,186
; FILING DATE: 21-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MORASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 23461-20007.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-670-186-2

```

```

Query Match          24.2%; Score 86.5; DB 2; Length 155;
Best Local Similarity 27.9%; Pred. No. 0.0012;
Matches 17; Conservative 15; Mismatches 28; Indels 1; Gaps 1;

```

```

QY 2 QQESQQLQCCNQVQKVRDE-CQCEAIKYIAEDQIQGQLHGESESVAVRAGEIVSSC 60
Db 83 QPRPRALRGCCNQVQKVRDE-CQCEAIKYIAEDQIQGQLHGESESVAVRAGEIVSSC 142

```

Oy 61 v 61
Db 143 i 143

RESULT 7

US-08-618-911-4
; Sequence 4, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig
; APPLICANT: Coughlan, Sean
; APPLICANT: Hu, David
; TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
; NUMBER OF INVENTIONS: SEEDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-618-911-4

Query Match

Best Local Similarity 23.4%; Score 83.5; DB 2; Length 158;
Matches 14; Conservative 23; Mismatches 19; Indels 3; Gaps 2;

Oy 2 QOESQOLQCCNOVKVRD-ECOCCEAIKTYAEDQIQGOLHGEESRVARAGEIVSSC 59

Db 88 EEEEGHMKCTEMSELSRSPKCKCKALOKIMENSE--ELEEKKKMKELINLATWC 144

RESULT 8

US-08-938-675A-2
; Sequence 2, Application US/08938675A
; Patent No. 6107287
; GENERAL INFORMATION:
; APPLICANT: de Lumen, Benito O.
; APPLICANT: Galvez, Alfredo F.
; TITLE OF INVENTION: Lunasin Peptides
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA

ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,675A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B98-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 158 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-938-675A-2

Query Match 23.4%; Score 83.5; DB 3; Length 158;
Best Local Similarity 23.7%; Pred. No. 0.0028;
Matches 14; Conservative 23; Mismatches 19; Indels 3; Gaps 2;

Oy 2 QOESQOLQCCNOVKVRD-ECOCCEAIKTYAEDQIQGOLHGEESRVARAGEIVSSC 59

Db 88 EEEEGHMKCTEMSELSRSPKCKCKALOKIMENSE--ELEEKKKMKELINLATWC 144

RESULT 9

US-08-991-300-2
; Sequence 2, Application US/08991300
; Patent No. 5973225
; GENERAL INFORMATION:
; APPLICANT: D'OVIDIO, RENATO
; APPLICANT: PORCEDDU, ENRICO
; APPLICANT: MERCHETELLI, CINZIA
; APPLICANT: CARDELLI, LUISA ERCOLI
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
; TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,300
; FILING DATE: 16-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IT MI 96/A 002663
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2264-0201-0X
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-300-2

Query Match 20.7%; Score 74; DB 2; Length 369;
Best Local Similarity 32.8%; Pred. No. 0.13;
Matches 21; Conservative 12; Mismatches 21; Indels 10; Gaps 3;

OY 2 QOESQQL--QCCNOKVRCDECCCEAIKYA-----EDQIQGQLHGESEVAQRAE 53
Db 235 QOSSHVWQCCQCOQLPQIPQSSRYEAIKIVSYIIQEQOVGSIQFQCCQ--PQQLG 292
OY 54 EIVS 57
Db 293 QCVS 296

RESULT 10
US-08-625-209A-2
Sequence 2, Application US/08625209A
Patent No. 5756671
GENERAL INFORMATION:
APPLICANT: Gyuris, Jeno
APPLICANT: Lamphere, Lou
APPLICANT: Diarella, Giulio
TITLE OF INVENTION: Cd37 Cell-Cycle Regulatory Protein,
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Eliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,209A
FILING DATE: 01-APR-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIU048.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 373 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-625-209A-2

Query Match 20.6%; Score 73.5; DB 1; Length 373;
Best Local Similarity 27.8%; Pred. No. 0.15;
Matches 20; Conservative 15; Mismatches 28; Indels 9; Gaps 2;
OY 2 QOESQQLQCCNOKVRCDECCCEAIKYAEDQIQG-----QLHGESEVAQRAEIV 56

Db 44 OKKEEIDRCGRCKRKVAECQ-----RKKELEVAEGSKAELERLQAEASTAAOGAGELG 99
OY 57 SSCGVRCMRQTR 68
Db 100 AEAGRCARRR 111

RESULT 11
US-07-955-905A-23
Sequence 23, Application US/07955905A
Patent No. 5770433
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
TITLE OF INVENTION: PRECURSOR
NUMBER OF SEQUENCES: 28
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/955,905A
FILING DATE: 21-JAN-1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 587 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum
FEATURE:
NAME/KEY: Protein
LOCATION: 1..587
OTHER INFORMATION: /note= "Vicilin from G. hirsutum"
US-07-955-905A-23

Query Match 19.9%; Score 71; DB 1; Length 587;
Best Local Similarity 26.5%; Pred. No. 0.54;
Matches 18; Conservative 15; Mismatches 29; Indels 6; Gaps 1;

OY 1 QOESQQLQCCNOKVRCDECCCEAIKYI-----AEDQIQGQLHGESEVAQRAE 54
Db 83 QRRYECCQECRQOEERQPCQOCRCIKRFEOQOQSOQFOECQOHQOEORPERKOQ 142
OY 55 IVSSCGVR 62
Db 143 CVAECRRER 150

RESULT 12
US-08-531-927B-2
Sequence 2, Application US/08531927B
Patent No. 5840491
GENERAL INFORMATION:
APPLICANT: Kakizuka, Akira
TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
Patent No. 5840491
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
114.625 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 Q00ESQ01QCCNQVKQVRD.....AQRAGETIVSCGVRCMRQTR 68

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	100.0	258	1	2S5_RICCO
2	125	35.0	295	1	P01089 riclinus com
3	110	30.8	164	1	2S55_HELAN
4	104.5	29.3	100	1	2S51_ARATH
5	104.5	29.3	104	1	2S54_CAPMA
6	103.5	28.0	104	1	2S53_CAPMA
7	102.5	28.7	130	1	2S51_CAPMA
8	101	28.3	164	1	ITRY_STINAR
9	98	27.5	80	1	P15459 arabidopsis
10	91	25.5	178	1	CG2L_LUPAN
11	91	25.5	180	1	2S52_BRANA
12	90.5	25.4	129	1	2S53_BRANA
13	89	24.9	170	1	ALIL_BRANU
14	88	24.6	172	1	2S52_ARATH
15	87	24.4	186	1	2S51_PICGL
16	86.5	24.2	155	1	2S52_CAPMA
17	86	24.1	127	1	2S52_BRANA
18	85.5	23.9	146	1	ALIL_STINL
19	85.5	23.9	319	1	2S5_BEREX
20	84.5	23.7	168	1	GDAS_WHEAT
21	84.5	23.7	166	1	PU1B_WHEAT
22	84	23.5	178	1	2S54_ARATH
23	83.5	23.4	110	1	2SSB_BRANA
24	83.5	23.4	158	1	2S51_BRANA
25	83	23.2	223	1	2S5_SOYB
26	82.5	23.1	289	1	GLU2_MAYE
27	82.5	23.1	289	1	GDAT_WHEAT
28	81.5	22.8	282	1	HOG3_HORVU
29	81.5	22.8	327	1	GDG3_WHEAT
30	79	22.1	304	1	GBD2_WHEAT
31	78.5	22.0	291	1	GBD1_WHEAT
32	76	21.3	220	1	GOBB_WHEAT
33	75.5	21.1	356	1	AVE3_AVE5A
					GLTA_WHEAT

34	75	21.0	307	1	GLTB_WHEAT	P10386 triticum ae
35	73	20.4	251	1	GD80_WHEAT	P08079 triticum ae
36	72.5	20.3	297	1	GD4_WHEAT	P04724 triticum ae
37	72.5	20.3	302	1	GD5_WHEAT	P21292 triticum ae
38	72	20.2	244	1	GD3_WHEAT	P04730 triticum ae
39	71.5	20.0	147	1	IAA_HORVU	P16969 hordeum vul
40	71.5	20.0	186	1	GL19_ORYSA	P29835 oryza sativ
41	71	19.9	305	1	HOG1_HORVU	P17990 hordeum vul
42	70.5	19.7	149	1	PRO7_ORYSA	P20698 oryza sativ
43	70	19.6	361	1	KOPF_SYNY3	P72830 synechocyst
44	69	19.3	143	1	IA16_WHEAT	P16159 triticum ae
45	69	19.3	291	1	GD42_WHEAT	P04722 triticum ae

ALIGNMENTS

RESULT 1
ID 2S5_RICCO STANDARD; PRT; 258 AA.
AC P01089;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RA MEDLINE=91109729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein
RT may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RA MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor
RT gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RA MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
RT protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RA MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
RT inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
RN [5]
RP FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC - SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY 2 DISULFIDE BONDS.
CC - PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CC CHAINS INVOLVE CYS-162 AND CYS-175.
CC - SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This swiss-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

```
CC or send an email to license@lsb.sib.ch).
CC -----
DR EMBL: X54158; CAA38097.1; -.
DR PIR: A01328; RZCS.
DR PIR: S11499; S11499.
DR PIR: S11500; S11500.
DR PIR: S11501; S11501.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; tryp_alpha_amy1_inh.
DR SMART: SM00499; AAI: 2.
DR SMART: SM00499; AAI: 2.
KW Seed storage protein; Signal; Allergen.
FT SIGNAL 1 21 PROBABLE.
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT PROPEP 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD_RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 MISSING (IN REF. 3).
FT CONFLICT 234 234 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
FT SEQUENCE 258 AA; 29290 MW; 27874CFC50EA1072 CRC64;

Query Match
Best Local Similarity 100.0%; Score 357; DB 1; Length 258;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 00ESQOLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESERVAORAGETVSSCG 60
DB 87 00ESQOLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESERVAORAGETVSSCG 146
QY 61 VRCKRQTR 68
DB 147 VRCKRQTR 154

RESULT 2
2SS5_HELAN STANDARD; PRT; 295 AA.
ID 2SS5_HELAN
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 162-173.
RC STRAIN=CV. GIANT GREY STRIPE;
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P.,
RA Nessler C.L., Thomas T.L.;
RT "Sequence and expression of a gene encoding an albumin storage
RT protein in sunflower."
RL Mol. Genet. 210:211-218(1987).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- PWM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
CC POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
CC PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
CC INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb.sib.ch/announce/
CC or send an email to license@lsb.sib.ch).
CC -----
DR EMBL: X06410; CAA29699.1; -.
DR PIR: S01062; S01062.
DR InterPro: IPR003612; AAI.
DR Pfam: PF00234; tryp_alpha_amy1_inh.
DR SMART: SM00499; AAI: 2.
DR SMART: SM00499; AAI: 2.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 161
FT CHAIN 162 295 2S SEED STORAGE PROTEIN.
FT SEQUENCE 295 AA; 34071 MW; 8958A106805142A1 CRC64;

Query Match
Best Local Similarity 35.0%; Score 125; DB 1; Length 295;
Matches 23; Conservative 22; Mismatches 19; Indels 10; Gaps 2;

QY 2 00ESQOLQCCNOVKYRDECCCEAIKTYIAEDQIQGQLHGSESERVAORAGETVSSCG 60
DB 74 RQEQQLQCCCELOINBGCCCEAVKVFREARQQVQGGGRLVPRRSGQDTQLKQK 133
QY 52 AGEIVSSGVRCMR 65
DB 134 AOLPNCVNLQSR 147

RESULT 3
2SS1_ARATH STANDARD; PRT; 164 AA.
ID 2SS1_ARATH
AC P15457;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein 1 precursor (2S albumin storage protein)
DE (NMW02-2S albumin 1).
GN AT2S1 OR AT4G27140 OR T24A18.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eusteroideae II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 38-73 AND 84-162.
RC STRAIN=CV. C24;
RA Kriebbers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
RA van Damme J., Segura M., Gheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family."
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Conceicao A.D.S., Kriebbers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Mambutt R., Murphy G., Volckaert G.,
RA Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermayer B., Maché R., Mueller M.,
RA Reichert B., Portelle D., Perez-Alonso M., Watson M., Schmidtheini T.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Gyromprez B., Chuang Y.-J., Vandebussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
```


RA Walzeberger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holtzer E., Brandt A., Peters S., van Stevenen M., Dijkse W.,
RA Moollman P., Klein lankhorst R., Rose M., Hauf J., Koester P.,
RA Benneiser S., Hempel S., Feldpausch M., Landwehr S., Van den Daele H.,
Ra De Keyser A., Blysshaert C., Glielen J., Villarroel S., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., Mcclay K., Mayes R.,
RA Pettett A., Rajendram M.A., Lyne M., Barnes V., Reehmann S.,
RA Borkova D., Bloeker H., Schaefe M., Grimm M., Lochner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fatmann B., Graderath K., Bauner D., Herzl A.,
RA Neumann S., Argitidou A., Vitale D., Ligouri R., Piravandi E.,
RA Massenot O., Guigley F., Glabaud G., Mendlein A., Felber R.,
RA Schnabel S., Hiller R., Schmidt W., Lechery A., Aboung S.,
RA Chedori F., Cooke R., Berger C., Monfort A., Casacubeta E.,
RA Gibbons T., Weber N., Vandembol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Furnelle B., Bent E., Johnson S., Tacón D., Jesse T.,
RA Heijnen D., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
RA Frishman D., Haese D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Devan M., Wilson R.R., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj I., Schutz K., Huang E., Spiegel L.,
RA Sehorn M., Murray J., Sheat P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Hatmon G., Edwards J.,
RA Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fillion L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spiehl J., Ryan E., Andrews S., Gelsel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonoli B., Zidanic M., Strong K., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shan R.,
RA Swaby I.K., O'Shaugnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohby N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RL Nature 402:769-777(1999).
CC -I- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -I- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -I- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
CC ARABIDOPSIS.
CC -I- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on way
CC use by non-profit institutions as long as its content is in no legal
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22032; AAA32743.1; '-'
DR EMBL; Z24745; CAAB0870.1; '-'
DR EMBL; AI3820; CAA01132.1; '-'
DR EMBL; AL035680; CAB8844.1; '-'
DR EMBL; AL161566; CAB79569.1; '-'
DR PIR; JA0161; NWMD1
DR PIR; S34676; S34676
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1rnp-amy1_lnh.
DR InterPro; IPR000617; Napln.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR PRINTS; PR00496; NAPLN.
DR Prodom; PD002498; Naplin; 1.
KW SMART; SKM0439; AAI; 1.
KW Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 37
FT CHAIN 38 73 2S SEED STORAGE PROTEIN 1 SMALL SUBUNIT.
FT PROPEP 74 83
FT CHAIN 84 162 2S SEED STORAGE PROTEIN 1 LARGE SUBUNIT.
FT PROPEP 163 164
SQ SEQUENCE 164 AA; 19014 MW; 2BF28CBA74D9832B CRC64;

```

Query Match 30.8%; Score 110; DB 1; Length 164;
Best Local Similarity 34.4%; Pred. No. 9,6e-05;
Matches 21; Conservative 12; Mismatches 28; Indels 0; Gaps 0;

QY 1 000ESQOLQCCNQYKQVNEDECCECAIKITAEQIQGQLHGEESEKVAQRAGEIYSSCG 60
      ||| | ||||| : : | : | : | : | : | : | : | : | : | : | : | : |
Db 87 000EQQLPQCCCNELRQEPEDCVCPFLTKQAKAVRLQGGHQMQRKVIQYTKALHPNVQD 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 v 61
Db 147 i 147

RESULT 4
2SS4_CAPMA STANDARD: PRT: 100 AA.
ID 2SS4_CAPMA
AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis maseikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Capparidaceae; Capparidaceae;
OX NCBI_TaxID=13395;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RA MEDLINE=94333405; PubMed=8055976;
RX Nirasawa S., Nishino T., Katahira M., Desugi S., Hu Z., Kurihara Y.;
  "Structures of heat-stable and unstable homologues of the sweet
  protein mabinlin. The difference in the heat stability is due to
  replacement of a single amino acid residue.";
  Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
  INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
  DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28 MABINLIN IV, A CHAIN.
FT NON_CONS 28 29
FT CHAIN 29 100 MABINLIN IV, B CHAIN.
FT DISULFID 4 49 BY SIMILARITY.
FT DISULFID 17 38 BY SIMILARITY.
FT DISULFID 39 87 BY SIMILARITY.
FT DISULFID 51 95 BY SIMILARITY.
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CMC64;

Query Match 29.3%; Score 104.5; DB 1; Length 100;
Best Local Similarity 31.3%; Pred. No. 0.00022;
Matches 21; Conservative 14; Mismatches 27; Indels 5; Gaps 2;

```



```

FT CHAIN 1 39 SMALL CHAIN.
FT NON CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R -> M.
FT VARIANT 53 57 MISSING (IN FORM II).
FT VARIANT 73 73 A -> S.
FT VARIANT 77 77 K -> R.
FT VARIANT 81 81 Q -> R.
FT VARIANT 87 87 H -> Q.
FT VARIANT 89 89 Q -> H.
FT VARIANT 91 91 Q -> Q.
FT VARIANT 97 97 E -> M.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
SQ SEQUENCE 130 AA; 14682 MW; EC02EA826D180DF2 CRC64;

Query Match 28.7%; Score 102.5; DB 1; Length 130;
Best Local Similarity 35.2%; Pred. No. 0.00045;
Matches 25; Conservative 10; Mismatches 25; Indels 11; Gaps 3;

QY 2 QGSEQDLOCCNQVRCQCECAIKYIA---EQOI-QGQLHGESE-----RVNQ 50
DB 44 QQPPPLQCCCHLHOEPLVCVPTLKAKAKKQOIQQGQGGQGLQHEIRIYQ 103
QY 51 RAGEIVSCGV 61
DB 104 TATHLPKVCNI 114

RESULT 8
2SS3_ARATH STANDARD; PRT; 164 AA.
AC P15459;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein 3 precursor (2S albumin storage protein)
DE (NMWU2-2S albumin 3).
GN AT2S3 OR AT4G37160 OR T24A18.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Kriebers E., Herdies L., de Clercq A., Seurlinck J., Leemans J.,
RA van Damme J., Segura M., Gheysen G., van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family.";
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Conceicao A.D.S., Kriebers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schneller C., Wambutt R., Murphy G., Voickaert G.,
RA Pohl T., Duesterhoeft A., Stiekema M., Ertlan K.-D., Terry N.,
RA Harris B., Ansoore W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermayer L., Macho R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Retelert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,

```

```

RA Vos P., Hobeisel J., Zimmermann W., Medler H., Ridley P.,
RA Langham S.A., McCullagh B., Billham L., Roben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-D., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Delfor E.,
RA Weltzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirse W.,
RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koettler P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth J., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., de Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
RA Borkova D., Bioecker H., Scharf M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiridou A., Vitale D., Iguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefder F., Cooke R., Berger C., Montfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Bedhia N., Gnoj L., Schütz K., Huang E., Spiegel L.,
RA Sektion M., Murray J., Sheel P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kallick J., Graves T., Harmon G., Edwards J.,
RA Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hiller L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
RN [4]
RP SEQUENCE OF 103-164 FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Raynal M., Grellet F., Laute M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -!- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M22035; AAA32745.1; -
CC EMBL; 224744; CAA80688.1; -
CC EMBL; AL035680; CAB38846.1; -
CC EMBL; AL161566; CAB79571.1; -
CC EMBL; Z17580; CAA79001.1; -
CC PIR; J0163; NMW03.
CC PIR; S34674; S34674.
CC InterPro; IPR003612; AAI.
CC InterPro; IPR001768; Cereal_1tryp_amyl_inh.
CC InterPro; IPR000617; Napin.
CC Pfam; PF00234; tryp_alpha_amyl; 1.
CC PRINTS; PR00496; Napin.
CC ProDom; PD002498; Napin; 1.
CC SMART; SM00499; AAI; 1.
CC Seed storage protein; Albumin; Signal; Multigene family.
KW

```

[illegible]

ID	2SS2_BRANA	STANDARD:	PRT:	178 AA.
AC	P01090.			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-AUG-1992 (Rel. 23, Last annotation update)			
DE	Napin 2 precursor (1.7S seed storage protein).			
OS	Brassica napus (Rape).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Brassica.			
OX	NCBI_TaxID=3708;			
RN	[1]			
RX	SEQUENCE FROM N.A.			
RX	MEDLINE=87308224; PubMed=3624251;			
RA	Josefsson L.-G., Lenman M., Ericsson M.L., Rask L.;			
RA	"Structure of a gene encoding the 1.7 S storage protein, napin, from			
RT	Brassica napus."			
RL	J. Biol. Chem. 262:12196-12201(1987).			
RN	[2]			
RP	REVISIONS.			
RA	Josefsson L.-G.;			
RL	submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87033665; PubMed=3771543;			
RA	Ericsson M.L., Roedin J., Lenman M., Glimelius K., Josefsson L.-G.,			
RA	Rask L.;			
RT	"Structure of the rapeseed 1.7 S storage protein, napin, and its			
RT	precursor."			
RL	J. Biol. Chem. 261:14576-14581(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. TOWER;			
RX	MEDLINE=84113267; PubMed=6689334;			
RA	Crouch M.L., Tenberge K.M., Simon A.E., Ferl R.;			
RT	"cDNA clones for Brassica napus seed storage proteins: evidence from			
RT	nucleotide sequence analysis that both subunits of napin are cleaved			
RT	from a precursor polypeptide."			
RL	J. Mol. Appl. Genet. 2:273-283(1983).			
CC	-1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE			
CC	TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING			
CC	ITS MATURATION.			
CC	-1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN			
CC	LINKED BY DISULFIDE BONDS.			
CC	-1- TISSUE SPECIFICITY: COYLEDONS AND THE AXIS.			
CC	-1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way com-			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; K01545; AAA33006.1; -			
DR	EMBL; J02586; AAA32997.1; -			
DR	EMBL; J02798; AAA87348.1; -			
DR	PIR; A01329; NMRP2			
DR	PIR; A29801; A29801.			
DR	PIR; A25997; A25997.			
DR	InterPro; IPR003612; AAI.			
DR	InterPro; IPR001768; Cereal_tryf-amy1_inh.			
DR	InterPro; IPR000617; Napin.			
DR	Pfam; PF00234; tryf_alpha_amy1. 1.			
DR	PRINTS; PR00496; NAFIN.			
DR	PRODom; PD002498; Napin. 1.			
DR	SMART; SM00499; AAI. 1.			
DR	Seed storage proteini; Signal; Multigene family.			
FT	SIGNAL	1	21	
FT	PROPEP	22	38	
FT	CHAIN	39	74	SMALL CHAIN.

DB 43 GQRPPLQCCNELHQEELVCVPTLKASKAVKQIROQGQGGQQLQHEISRIYQ 102
 QY 51 RAGEIVSSCGV 61
 DB 103 TATHLPVNCNI 113

RESULT 13
 2SS2_ARATH STANDARD; PRT; 170 AA.
 AC P15458;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 2S seed storage protein 2 precursor (2S albumin storage protein)
 DE (NMWU2-2S albumin 2)
 GN AT282 OR AT4627150 OR T24A18.100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. C24;
 RA Kriebers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
 RA van Damme J., Segura M., Ghysen G., van Montagu M.,
 RA Vandekerckhove J.;
 RT "Determination of the processing sites of an Arabidopsis 2S albumin
 RT and characterization of the complete gene family."
 RL Plant Physiol. 87:859-866(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. C24;
 RA Conceicao A.D.S., Kriebers E.;
 RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RX MEDLIN=20083488; PubMed=10617198;
 RX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansgore W., Brandt P., Grivell L.A., Rieger M.,
 RA Welcheselgatter M., de Simone V., Obermaier B., Maehle R., Mueller M.,
 RA Kreits M., Delseny M., Pulgomech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnsels J., Zimmermann W., Medler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bihlman L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbusche F.,
 RA Breken M., Weltens I., Voet M., Bastlaens I., Aert K., Defoor E.,
 RA Weltzenegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetler P.,
 RA Benneker S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA de Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K.,
 RA Petlett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borikova D., Bloeker H., Scharfe M., Grimm M., Loehnert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Hertz A.,
 RA Neumann S., Argitlou A., Vitale D., Liguori R., Pitravanti E.,
 RA Massenet O., Quigley F., Clabaud G., Mendenhall A., Felder R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Schiffr D., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Baigues M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Fristman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Beyan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedina N., Gnoj L., Schutz K., Huang B., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes K., Abu-Threihel J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,

RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton L., Mardis E., Dante M., Pepin K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hiller L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Yil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Gill S.,
 RA Grant S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martensen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 RN [4]
 RP SEQUENCE OF 1-90 AND 118-170 FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Raynal M., Grellert F., Lande M., Meyer Y., Cooke R., Delseny M.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as there is no content in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M22034; AAA32744.1; -;
 DR EMBL: 224745; CAA80871.1; -;
 DR EMBL: ALO35680; CAB38845.1; -;
 DR EMBL: AL165565; CAB79570.1; -;
 DR EMBL: Z17598; CAA79010.1; -;
 DR EMBL: Z17594; CAA79008.1; -;
 DR PIR: JAO162; NMW02.
 DR PIR: S34677; S34677.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_aml_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryp_alpha_aml_1.
 DR PRINTS: PR00496; NAPIN.
 DR PRODOM: PD002498; Napin. 1.
 DR SMART: SM00499; AAI. 1.
 KV Seed storage protein: Albumin; Signal; Multigene family.
 FT SIGNAL 1 21
 FT PROPEP 22 37
 FT CHAIN 38 72 2S SEED STORAGE PROTEIN 2 SMALL SUBUNIT
 FT (BY SIMILARITY).
 FT PROPEP 73 88 2S SEED STORAGE PROTEIN 2 LARGE SUBUNIT
 FT CHAIN 89 170 (BY SIMILARITY).
 FT SEQUENCE 170 AA; 19361 MW; 0A0562D5FAB56275 CRC64;
 SQ

Query Match 24.9%; Score 89; DB 1; Length 170;
 Best Local Similarity 32.4%; Pred. No. 0.014;
 Matches 23; Conservative 14; Mismatches 28; Indels 6; Gaps 3;

QY 2 QGSGQLQCCNQVKNVDECCCEAIKIYADQIQGGHGE-ESRYAQRAGEIVSSCG 60
 DB 93 QGSHQILQCCSELHQEELVCVPTLKASKAVKQIROQGQGGQQLQHEISRIYQ 151

QY 61 VR---CMKQRT 67
 DB 152 IQQVGECPPT 162

RESULT 14
 2SSL_PICGL STANDARD; PRT; 172 AA.
 ID 2SSL_PICGL

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:36 : Search time 77.42 Seconds
(without alignments)
151.946 Million cell updates/sec

Title: US-09-913-351-3
Perfect score: 357
Sequence: 1 QQESQQLQCCNQVKNQVRD.....AQRAGETVSSGCVACMQR 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriaph:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	39.5	139	10	P93198
2	132.5	37.1	323	10	Q39928
3	128	35.9	153	10	Q9AUD1
4	124	34.7	65	10	Q9S872
5	122	34.2	66	10	Q9S870
6	122	34.2	141	10	Q39649
7	112	31.4	64	10	Q3AVK2
8	109	30.5	161	10	Q81410
9	107.5	30.1	165	10	Q40850
10	102	28.6	153	10	Q39235
11	99	27.7	172	10	Q40997
12	97	27.2	167	10	Q81411
13	96	26.3	184	10	Q42444
14	95.5	26.8	91	10	Q9S9E5
15	95.5	26.8	162	10	Q64929
16	95	26.6	162	10	Q40998

17	92	25.8	165	10	Q9FH31
18	91	25.5	86	10	Q9S9F0
19	91	25.5	88	10	Q9S9E7
20	91	25.5	125	10	P80208
21	91	25.5	178	10	Q42413
22	91	25.5	178	10	Q42473
23	91	25.5	178	10	Q96339
24	91	25.5	180	10	Q42469
25	90.5	25.4	91	10	Q9S9E5
26	90.5	25.4	155	10	Q9ZN24
27	90	25.2	173	10	Q64930
28	89.5	25.1	148	10	Q9XHP1
29	89	24.9	162	10	Q64932
30	89	24.9	190	10	Q40995
31	88.5	24.8	173	10	Q9Z8R1
32	88	24.6	139	10	Q39795
33	88	24.6	285	10	Q41603
34	87.5	24.5	152	10	Q41168
35	87.5	24.5	156	10	Q941R0
36	87	24.4	88	10	Q9S9E9
37	87	24.4	88	10	Q9S9E8
38	86	24.1	178	10	Q39344
39	85.5	23.9	146	10	Q9JRC2
40	85.5	23.9	170	10	Q64931
41	85.5	23.9	174	10	Q41169
42	85.5	23.9	318	10	Q41545
43	85	23.8	155	10	Q41167
44	84.5	23.7	173	10	Q81412
45	84	23.5	178	10	Q42490

ALIGNMENTS

RESULT 1
P93198 PRELIMINARY: PRT: 139 AA.
AC P93198;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND;
RA Teuber S.S., Dandekar A.M., Peterson W.R.;
RT "Juglans regia 2S albumin seed storage protein precursor."
RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U66866; AAB41308.1; -.
DR InterPro: IPR001768; Cereal_1tryp_1.
DR InterPro: IPR003612; AAT.
DR InterPro: IPR001768; Cereal_1tryp_1.
DR InterPro: IPR004480; Glutelin.
DR Pfam: PF00234; tryp_alpha_1.
DR PRINTS: PR00211; GLUTELIN.
DR SMART: SM00499; AAT; 1.
FT NON_TER
SQ SEQUENCE 139 AA: 16373 MW: 02D0E55E67164F23 CRC64;

Query Match 39.5%; Score 141; DB 10; Length 139;
Best Local Similarity 35.4%; Pred. No. 1.9e-08;
Matches 23; Conservative 15; Mismatches 27; Indels 0; Gaps 0;

QY 1 QQESQQLQCCNQVKNQVRDCECEAIKTYAEQIQGGQLHGESRVAQRAGEIYSSCG 60
DB 67 EDNQRHFRCQCOOLSSQMBDCCEGLRVRRQDQDQGLRGEEMEVQASADLPNECG 126
QY 61 VRQMR 65

Db 127 ISSOR 131

RESULT 2

ID 039928 PRELIMINARY; PRT: 323 AA.

AC 039928;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE 2S ALBUMIN PRECURSOR (FRAGMENT).

OS Helianthus annuus (Common sunflower).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;

OC Heliantheae; Helianthus.

OX NCBI_TaxID=4232;

RM [1]

RP SEQUENCE FROM N.A.

RC TISSUE=FLOWERING SEED;

RA Thoyts P.J.E., Millichip M., Stobart A.K., Griffiths W.T.,

RA Napier J.A., Shewry P.R.;

RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.

DR EMBL: X76101; CAA53710.1; .

DR InterPro: IPR003612; AAI.

DR InterPro: IPR001768; Cereal_1ryp_aml_1nh.

DR InterPro: IPR000617; Napin.

DR Pfam: PF00234; 1ryp_alpha_aml; 2.

DR PRINTS: PR00496; NAPIN.

DR SMART: SM00499; AAI; 2.

KM Signal.

FT SIGNAL

FT CHAIN

SQ SEQUENCE 323 AA; 38282 MW; A15378AFAD52D6B3 CRC64;

Query Match 37.1%; Score 132.5; DB 10; Length 323;
Best Local Similarity 40.6%; Pred. No. 4.2e-07;

Matches 26; Conservative 14; Mismatches 19; Indels 5; Gaps 1;

OY 1 QOSESQQLQCCNQYKVRDECCEAIKTYA-----EDQIQGQLHGESESRVAGRAGEI 55

Db 249 ROEEQGGQGCCNELQNVRECCCEAIKEVGQRMQOQOORHQYGGQQTVERILENL 308

OY 56 VSSC 59

Db 309 PNOG 312

RESULT 3
09AUD1 PRELIMINARY; PRT: 153 AA.

AC 09AUD1;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE 2S ALBUMIN.

OS Sesamum indicum (Oriental sesame) (Gingelly).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Lamiales; Pedaliaceae; Sesamum.

OX NCBI_TaxID=4182;

RN [1]

RP SEQUENCE FROM N.A.

RA Tal S.S.K., Tzen J.T.C.;

RT "Molecular cloning of three storage proteins in sesame."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF240005; AAK15088.1; .

DR HSSP: P01085; IHSS.

DR InterPro: IPR003612; AAI.

DR InterPro: IPR001768; Cereal_1ryp_aml_1nh.

DR InterPro: IPR000480; Glutelin.

DR InterPro: IPR000617; Napin.

DR Pfam: PF00234; 1ryp_alpha_aml; 1.

DR PRINTS: PR00211; GLUTELIN.

DR PRINTS: PR00496; NAPIN.

DR SMART: SM00499; AAI; 1.

SQ SEQUENCE 153 AA; 17504 MW; 6A8F9117DAE81568 CRC64;

Query Match 35.9%; Score 128; DB 10; Length 153;
Best Local Similarity 37.7%; Pred. No. 6.1e-07;

Matches 23; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

OY 2 QOSESQQLQCCNQYKVRDECCEAIKTYAEDQIQGQLHGESESRVAGRAGEIVSSCGV 61

Db 82 QOSESQQLRCCQQLRVNDRCEAIKTYAVRQOQEGVQOESQOQYGRADLPKRCWM 141

OY 62 R 62

Db 142 R 142

RESULT 4

ID 09S872 PRELIMINARY; PRT: 65 AA.

AC 09S872;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE 7.3 KDA NAPIN-LIKE PROTEIN LARGE CHAIN (FRAGMENT).

OS Ricinus communis (Castor bean).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.

OX NCBI_TaxID=3988;

RN [1]

RP SEQUENCE.

RX MEDLINE=97135090; PubMed=8980648;

RA Neumann G.M., Condron R., Polya G.M.;

RT "Purification and sequencing of napin-like protein small and large

RT chains from Momordica charantia and Ricinus communis seeds and

RT determination of sites phosphorylated by plant Ca(2+)-dependent

RT protein kinase."

RL Biochim. Biophys. Acta 1298:223-240(1996).

DR InterPro: IPR003612; AAI.

DR InterPro: IPR000617; Napin.

DR PRINTS: PR00496; NAPIN.

DR SMART: SM00499; AAI; 1.

SQ SEQUENCE 65 AA; 7350 MW; CB9264C4C917F498 CRC64;

Query Match 34.7%; Score 124; DB 10; Length 65;
Best Local Similarity 36.2%; Pred. No. 6.8e-07;

Matches 21; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

OY 4 ESOQLQCCNQYKVRDECCEAIKTYAEDQIQGQLHGESESRVAGRAGEIVSSCGV 61

Db 1 EERSLRGCCDHLKQWQSCRCBGLRQATEQQSQQLQGVDFEAFRTAANLPMSGCV 58

RESULT 5

ID 09S870 PRELIMINARY; PRT: 66 AA.

AC 09S870;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)

DE 7.9 KDA NAPIN-LIKE PROTEIN LARGE CHAIN (FRAGMENT).

OS Momordica charantia (Bitter melon) (Balsam pear).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Cucurbitales; Cucurbitaceae; Momordica.

OX NCBI_TaxID=3673;


```

Query Match          30.5%; Score 109; DB 10; Length 161;
Best Local Similarity 27.2%; Pred. No.8.7e-05;
Matches 22; Conservative 21; Mismatches 20; Indels 18; Gaps 2
QY      1 000ESQOLQOCCNOVKOVRDECCEAIKRYIADQI-----00GOLHGEESR 47
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      66 ERRRNQPSERCCEEIQRMSPCRCQAIGDTLEDVPMDSQGAPLNRGRRGMEERE 125
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      48 VAORAGEIVSSCGV-----RC 63
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      126 VLRRAEELPNTCNVPQSPRC 146
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT    9
Q040850   PRELIMINARY; PRT; 165 AA.
AC Q040850;
ID Q040850;
DR 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN SEED STORAGE PROTEIN.
OS EMB25.
OC Picea glauca (White spruce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=33390;
[1]
RN SEQUENCE FROM N.A.
RP Dong J.-Z., Dunstan D.I.;
RL "Gene expression during somatic embryogenesis.";
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L47745; AAB01560.1; -.
DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR SMART; SM00494; AAT; 1.
SQ SEQUENCE 165 AA; 19396 MW; 69CD273982C40D20 CRC64;

Query Match          30.1%; Score 107.5; DB 10; Length 165;
Best Local Similarity 26.6%; Pred. No.0.00013;
Matches 21; Conservative 20; Mismatches 21; Indels 17; Gaps 1
QY      1 000ESQOLQOCCNOVKOVRDECCEAIKRYIAD-----QIQGQLNGE 43
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      66 ERPDQPSERCCCEEIQRMSPCRCQAIGDTLEDVPMDSQGAPLNRGRRGGRGM 125
       |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      44 ESERYAORAGEIVSSCGVR 62
       |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      126 EEVEVRRRAEELPNTCNVR 144
       |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT    10
Q099235   PRELIMINARY; PRT; 153 AA.
AC Q099235;
ID Q099235;
DR 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN CONGLUTIN DELTA.
OS Lupinus angustifolius (Narrow-leaved blue lupine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Genistaeae; Lupinus.
OX NCBI_TaxID=3871;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-UNICROP; TISSUE=COTYLEDON;
RX MEDLINE=91355912; Pubmed=2103479;
RA Gayler K.R., Johanson S., Macfarlane A.J., Milley G.G., Baldi M.,
RA Blagrove R.J., Johnson E.D.;
```

[illegible]

Best Local Similarity 25.6%; Pred. No. 0.0028;
Matches 21; Conservative 21; Mismatches 21; Indels 19; Gaps 4;

QY 1 000ESQQLQCCCNQVQVDECCCEAIKYI-----AEDQIQ--GQLHG---EESF 46

Db 66 ERRREOPSSCCNELERMSPOCRPAIQOVLDOSSAFMDESDALNQRGRREGRRREQ 125

QY 47 RVAQRAGEIVSSCGV-----RC 63

Db 126 EMARAAVLPDTCNVQESPRC 147

Search completed: September 9, 2002, 12:45:37
Job time: 294 sec

PT Ricinus communis -
XX
PS Claim 3; Page 30; 38pp; Spanish.
XX
CC The invention relates to a novel glycoconjugate (A) comprising a
CC 50-250 kDa polysaccharide (PS), containing one phosphate group per
CC 5-25 monosaccharide units, and a polypeptide (PP) having a particular
CC consensus sequence. In PS, at least 40% of the monosaccharide residues
CC are mannose and the remainder are glucose and/or galactose. The main
CC chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has
CC the consensus sequence:
CC zaa2bdc(QEKK)2(Z') (LIVM)2CC(CZ') (QEH) (LV)2SC2C222dGZCZ(VIIM)2ZICZg
CC where Z' = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;
CC a = 3-48; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;
CC the parentheses indicate the preferred sequence. This sequence
CC corresponds to amino acids 157-190 of the Ricinus communis (castor bean)
CC 2S albumin storage protein which is used to generate the glycoconjugate
CC of the invention, with a polysaccharide from *Candida utilis*. (A) modifies
CC the immune response in humans and animals, especially it reduces the
CC amount of tumour necrosis factor (TNF) produced and stimulates the
CC mononuclear-phagocyte system and expands the granulocyte-macrophage
CC compartment. (A) are used for treating disorders of the immune system,
CC e.g. infections and tumours. (A) do not interfere with metabolic
CC processes in the liver and are active when given orally. PP significantly
CC increases the antigenic response to weakly immunogenic PS, including
CC induction of a T-dependent response.
CC
XX
SQ Sequence 34 AA;

Query Match 100.0%; Score 183; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 6.8e-18;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PSQSGCGRGIOEQNLRCQCEYIKQVSGGPRR 34
Db 1 psggcrqgqigegqnlrqqceyikqvgsggprtr 34
|||||
1 psggcrqgqigegqnlrqqceyikqvgsggprtr 34

RESULT 2
AAB85946
ID AAB85946 standard; peptide: 17 AA.
XX
AC AAB85946;
XX
DT 30-NOV-2001 (first entry)
XX
DE Castor bean 2S small subunit protein fragment.
XX
KW Napin gene; promoter; Brassica; phenotype; transcription; EA9; 3H11;
KW 2A11; alpha-amylase.
XX
OS Ricinus communis.
XX
PN US6281410-B1.
XX
PD 28-AUG-2001.
XX
PF 15-JAN-1999; 99US-0232861.
XX
PR 29-APR-1988; 88US-0188361.
PR 02-NOV-1988; 88US-0267685.
PR 10-AUG-1993; 93US-0105852.
PR 07-JUN-1995; 95US-0484941.
PR 07-MAR-1997; 97US-0812665.
PR 31-JUL-1986; 86US-0891529.
PR 26-MAY-1987; 87US-0054369.
PR 28-JUL-1987; 87US-0078538.
PR 25-JAN-1988; 88US-0147781.
PR 15-MAR-1988; 88US-0168190.
XX
PA (CALJ) CALGENE LLC.
XX

PI Knauf VC, Kridl JC;
XX
DR WPI; 2001-564354/63.
XX
XX
PT Obtaining a plant that produces a seed with a modified phenotype or
PT altering a seed phenotype, comprises transforming a plant cell with a
PT DNA construct consisting of operably linked components in the direction
PT of transcription -
XX
PS Example 9; Fig 6; 68pp; English.
XX
CC The invention provides a method for obtaining a plant which produces
CC at least one seed having a modified phenotype. The method involves
CC transforming a host plant cell with a DNA construct which consists of
CC operably linked components in the direction of transcription, a promoter
CC region from a Brassica plant gene, a DNA sequence of interest other than
CC the native coding sequence, and a transcription termination region. The
CC method is useful for obtaining plants having modified phenotype or for
CC altering the phenotype of a plant seed or tissue. The DNA constructs are
CC used in manipulating plant cells to provide for regulated transcription,
CC such as light inducible transcription, in a plant tissue or plant part of
CC interest at particular stages of plant growth or in response to external
CC control. These constructs are also used for modulation of expression of
CC endogenous products as well as production of exogenous products in the
CC seed. Sequences AAB85941-47 represent fragments of storage proteins used
CC in comparison studies with the storage protein 2A11.
CC
XX
SQ Sequence 17 AA;

Query Match 42.9%; Score 78.5; DB 22; Length 17;
Best Local Similarity 94.4%; Pred. No. 0.00052;
Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 13 QONLRQCEYIKQVSGQ 30
Db 1 qgnlrqqceyl-qgyvsq 17
|||||
1 qgnlrqqceyl-qgyvsq 17

RESULT 3
AAR62762
ID AAR62762 standard; Protein: 110 AA.
XX
AC AAR62762;
XX
DT 07-JUL-1995 (first entry)
XX
DE Synthetic 12 kD precursor protein of Brazil nut.
XX
KW Stabilisation; toxic; peroxisome; yeast; food additive; brazil nut.
KW Synthetic.
XX
OS Synthetic.
XX
PN WO9424289-A.
XX
PD 27-OCT-1994.
XX
PF 19-APR-1994; 94WO-FR00438.
XX
PR 19-APR-1993; 93FR-0004583.
XX
PA (EURO-) EUROLYSTINE.
XX
PI Ito H, Iabat N, Nicaud J, Pardo D, Raynal A, Sugimoto S;
XX
DR WPI; 1994-341868/42.
DR N-PSDB; AAQ73069.
XX
PT Improving stability of protein or peptide in yeast - by
PT expressing with a peroxisomal targeting sequence, allowing
PT accumulation of unstable or toxic products, and new yeast strains
PT useful as food additives
XX


```

AAM86271
ID AAM86271 standard; Protein; 111 AA.
XX
AC AAM86271;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene BN153KW amino acid sequence.
XX
KW Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN19; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
OS Synthetic.
XX
FH Key
FH Misc-difference 4 Location/Qualifiers
FT /label= E4K
FT /note= "wild-type Glu is replaced with Lys"
FT Misc-difference 6
FT /label= R6K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 11
FT /label= R11K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 16
FT /label= S16K
FT /note= "wild-type Ser is replaced with Lys"
FT Misc-difference 19
FT /label= R19K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 23
FT /label= R23K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 36
FT /label= R36K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 37
FT /label= R37K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 41
FT /label= P41K
FT /note= "wild-type Pro is replaced with Lys"
FT Misc-difference 42
FT /label= W42H
FT /note= "wild-type Trp is replaced with His"
FT Misc-difference 58
FT /label= R58K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 63
FT /label= R63K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 68
FT /label= R68K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 77
FT /label= R77K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 82
FT /label= R82K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 83
FT /label= R83K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 86
FT /label= R86K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 94
FT /label= R94K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 101
FT /label= R101K

```

```

FT XX /note= "wild-type Arg is replaced with Lys"
PN MO9845458-A1.
XX
XX 15-OCT-1998.
PD
XX
PF 06-APR-1998; 98MO-US06673.
XX
PR 08-APR-1997; 97US-0042827.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Gutteridge S;
XX
XX WPI; 1998-609902/51.
DR N-PSDB; AAW80251.
XX
XX New modified Brazil Nut 2S albumin proteins - useful to provide
PT transgenic seed crops whose proteins have increased levels of
PT essential amino acids
XX
XX Claim 6; Fig 8; 80pp; English.
PS
XX
XX The invention relates to modified Brazil Nut 2S albumin seed storage
CC proteins and nucleic acids encoding the modified proteins. The modified
CC Brazil nut 2S albumin proteins are rich in essential amino acids and the
CC encoding genes selected from a group comprising of BNCNSS, BN11, BN15,
CC BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes.
CC Chimeric genes comprising a nucleic acid fragment encoding a modified
CC Brazil nut 2S protein operably linked to suitable regulatory sequences,
CC integrated into the genome of a transformed plant can be used in
CC methods for increasing the essential amino acid content of seeds. The
CC products can be used to provide a human and animal dietary protein in
CC seed crop plants that is relatively high in essential amino acid content.
CC The present sequence represents the amino acid sequence of a modified
CC Brazil Nut 2S albumin gene BN153KW.
XX
SQ Sequence 111 AA:

Query Match 37.2%; Score 68; DB 19; Length 111;
Best Local Similarity 44.0%; Pred. No. 0.11;
Matches 11; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 3 QGCGRCGQIQEQRNRCQERYIKQOV 27
DB 2 gekckeqmqkqkmlkcnkymkqgm 26

RESULT 9
AAM86265
ID AAM86265 standard; Protein; 111 AA.
XX
AC AAM86265;
XX
DT 01-MAR-1999 (first entry)
XX
DE Modified Brazil nut 2S albumin gene BNCNSS amino acid sequence.
XX
KW Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN19; AT2S1BN153W;
KW chimeric; dietary protein; essential amino acid.
XX
OS Bertholletia excelsa.
OS Synthetic.
XX
FH Key
FH Misc-difference 37 Location/Qualifiers
FT /label= R37K
FT /note= "wild-type Arg is replaced with Lys"
FT Misc-difference 58
FT /label= R58K
FT /note= "wild-type Arg is replaced with Lys"

```


THIS PAGE BLANK (USPTO)

DB 38 NQPLCRPOFQOHQHLRACQRYIRRAQRG 67

RESULT 2

US-08-453-924-3

Sequence 3, Application US/08453924

Patent No. 5608152

GENERAL INFORMATION:

APPLICANT: Kridl, Jean C.

APPLICANT: Knauf, Vlc C.

TITLE OF INVENTION: Seed-specific Transcriptional Regulation

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Well, Gotshal & Manges

STREET: 2882 Sand Hill Road, Ste. 280

CITY: Menlo Park

STATE: CA

COUNTRY: U.S.A.

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,924

FILING DATE: 30-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/742,834

FILING DATE: 08-AUGUST-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/550,804

FILING DATE: 09-JULY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/147,781

FILING DATE: 25-JANUARY-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/078,538

FILING DATE: 28-JULY-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 06/891,529

FILING DATE: 31-JULY-1986

ATTORNEY/AGENT INFORMATION:

NAME: Rae-Ventler, Barbara

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: CGNE-037/04US

TELEPHONE: (415) 926-6200

TELEFAX: (415) 854-3713

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 184 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-453-924-3

Query Match 37.2%; Score 68; DB 1; Length 184;

Best Local Similarity 41.4%; Pred. No. 0.071;

Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

DB 47 CRKEFOQAQHLKACQWHLKQAMQSGSG 75

US-08-670-186-6

Sequence 6, Application US/08670186

Patent No. 5859343

GENERAL INFORMATION:

APPLICANT: SUN, SAMUEL S.M.

APPLICANT: XIONG, LIWEN

APPLICANT: HU, ZHONG

APPLICANT: CHEN, HANG

TITLE OF INVENTION: RECOMBINANT SWEET PROTEIN MABINLIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 PENNSYLVANIA AVE NW, STE. 5500

CITY: WASHINGTON

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/670,186

FILING DATE: 21-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 23461-20007,00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 822-0168

TELEX: 90-4030 MRSNFOERSWSH

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 158 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-670-186-6

Query Match 35.0%; Score 64; DB 2; Length 158;

Best Local Similarity 36.7%; Pred. No. 0.2;

Matches 11; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

DB 38 NQPLCRPOFQOHQHLRACQRYIRRAQRG 67

US-08-618-911-4

Sequence 4, Application US/08618911

Patent No. 5850016

GENERAL INFORMATION:

APPLICANT: Jung, Rudolf

APPLICANT: Hastings, Craig

APPLICANT: Coughlan, Sean

APPLICANT: Hu, David

TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.

STREET: 700 Capital Square, 400 Locust Street

CITY: Des Moines

STATE: Iowa

COUNTRY: USA

ZIP: 50309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

TELEPHONE: 617-248-7000
 TELEFAX: 617-248-7100
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 542 amino acids
 TYPE: amino acid
 TOPOLOGY: 1 linear
 MOLECULE TYPE: peptide
 ORGANISM SOURCE:
 ORGANISM: Saccharomyces cerevisiae
 IMMEDIATE SOURCE:
 TISSUE: fractional yeast SSNP (fySSNP) (predicted)
 US-08-338-809-8

Query Match 26.2%; Score 48; DB 1; Length 542;
 Basic Local Similarity 24.0%; Pred. No. 92;
 Matched 12; Conservative 8; Mismatches 6; Indels 24; Gaps 1;
 Q: QIQEOONEROCCEYIKO-----QVSGGPPR 34
 I: I I I I I
 D: Q339 QIQEOONEROCCEYIKO-----QVSGGPPR 308

Search completed: September 9, 2002, 12:44:13
 Job time: 235 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:43:30 ; Search time 45.25 Seconds
(without alignments)
72.200 Million cell updates/sec

Title: US-09-913-351-4

Perfect score: 183

Sequence: 1 PSQGGCRGQIQEQGNLRQCEYIKQVSGGPRR 34

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR_71:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	258	1 RZCS	2S seed storage pr
2	82	44.8	106	2 S26636	napin n1d - rape
3	80	43.7	110	2 S20350	napin n1d - rape
4	76.5	41.8	154	2 S14947	2S albumin - Brazi
5	74	40.4	139	2 T09878	albumin 2S storage
6	73	39.9	32	2 B59346	seed storage prote
7	71	38.8	28	2 A25802	2S seed storage pr
8	71	38.8	37	2 S35581	major allergen sma
9	71	38.8	39	2 S01791	allergen Sin a I s
10	71	38.8	127	2 S65447	allergen Sin a I s
11	71	38.8	145	2 PC1246	Sin a I allergen 1
12	71	38.8	145	2 S65478	allergen Sin a I
13	71	38.8	145	2 S65481	allergen Sin a I
14	71	38.8	145	2 S65482	allergen Sin a I
15	71	38.8	145	2 S65480	allergen Sin a I
16	71	38.8	145	2 S65479	allergen Sin a I
17	71	38.8	152	2 PS0427	allergen Sin a I
18	71	38.8	174	2 PS0425	napin AH1 precursor
19	71	38.8	180	2 S52025	napin BA3 precursor
20	70	38.3	139	2 T09850	napin (clones BMN
21	70	38.3	146	2 S14946	albumin 2S storage
22	69	37.7	104	2 S48176	2S seed storage pr
23	69	37.7	124	2 A37931	mablinin I-1 - Yun
24	69	37.7	145	2 PC1247	napin - Swedish tu
25	69	37.7	178	1 NMRP2	Sin a I allergen 2
26	69	37.7	178	2 S07838	napin 2 precursor
27	69	37.7	178	2 A25957	napin B - rape
28	69	37.7	178	2 S25137	napin precursor (n
29	68	37.2	178	2 S15382	2S storage protein
					napin napb - rape

30	68	37.2	178	2 S25127	2S storage protein
31	68	37.2	180	2 S10018	napin (clone BngNA
32	67.5	36.9	37	2 A23617	conglutin delta-2
33	67	36.6	39	2 JQ2321	calmodulin antagon
34	67	36.6	104	2 S48178	mablinin IIR - Yun
35	67	36.6	178	2 S25134	2S storage protein
36	66.5	36.3	100	2 S48180	mablinin IV - Yun
37	66.5	36.3	153	2 A33090	conglutin delta pr
38	66	36.1	33	2 P00801	calmodulin antagon
39	66	36.1	133	2 NMRP1	napin 1 precursor
40	66	36.1	164	1 NMRM1	2S albumin 1 precu
41	66	36.1	186	2 A29802	napin precursor (9
42	65	35.5	39	2 P00802	calmodulin antagon
43	64.5	35.2	166	1 NMRM4	2S albumin 4 precu
44	64	35.0	37	2 S16183	napin n1ii - rape
45	63.5	34.7	164	1 NMRM3	2S albumin 3 precu

ALIGNMENTS

RESULT 1

RZCS

2S seed storage protein precursor - castor bean

N:Alternate names: 2S albumin precursor

C:Species: Ricinus communis (castor bean)

C:Date: 14-Nov-1993 #sequence_revision 08-Feb-1996 #text_change 18-Jun-1999

C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222

R:Rivin, S.D.; Lord, J.M.

Nucleic Acids Res. 18, 5890, 1990

A:Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.

A:Reference number: S11499; MUID:91016940

A:Accession: S11499

A:Molecule type: DNA

A:Residues: 1-258 <IRW>

A:Cross-references: EMBL:X54158; NID:Q21067; PIDN:CAA38097.1; PID:Q21068

A:Note: The authors translated the codon CTC for residue 14 as Phe, CCA for residue 7

R:Rivin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.

Mol. Gen. Genet. 222, 400-408, 1990

A:Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro

A:Reference number: S11500; MUID:91109729

A:Accession: S11500

A:Molecule type: mRNA

A:Residues: 1-13, 'F', 'L', 'S', 'T', '75-258 <IR2>

A:Experimental source: clone 1494

A:Accession: S11501

A:Molecule type: mRNA

A:Residues: 'M', '4', 'L', 'S', '7-13', 'F', '15-21 <IRF>

A:Experimental source: clone 10a12

A:Accession: S27221

A:Molecule type: protein

A:Residues: 'X', '37', 'X', '39-45', 'X', '158-161', 'X', '163-174', 'X' <IRX>

R:Sharief, F.S.; Li, S.S.L.

J. Biol. Chem. 257, 14753-14759, 1982

A:Title: Amino acid sequence of small and large subunits of seed storage protein from

A:Reference number: A92357; MUID:83082772

A:Accession: A01328

A:Molecule type: protein

A:Residues: 157-190, 194-221, 'Q', '223-225, 230-233, 'N', '235-254, 'Q', '256-258 <SHA>

A:Note: 230-Set was also found

A:Note: There is considerable similarity between residues 181-231 of this protein and

R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.

Biochem. J. 213, 543-545, 1983

A:Title: Structural relationship between barley (Hordeum vulgare) trypsin inhibitor a

A:Reference number: A90322; MUID:83308577

A:Contents: annotation

A:Note: this protein is homologous with trypsin inhibitor from barley

C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and

C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:2-35/Domain: propeptide #status predicted <PRO>
F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC>
F:87-156/Product: probable 2S seed storage protein small chain 2 #status predicted <LCHI>
F:157-190/Product: 2S seed storage protein small chain #status experimental <SML>
F:194-258/Product: 2S seed storage protein large chain #status experimental <LRG>
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 183; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 4.4e-17;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSQGRGQIQEQNLRCQCEYIKQVSGGPRR 34
Db 157 PSQGRGQIQEQNLRCQCEYIKQVSGGPRR 190

RESULT 2
S26636

napin n1b - rape

C:Species: Brassica napus (rape)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 04-Oct-1996

C:Accession: S26636

R:Monsalve, R.I.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R.

FEBS Lett. 295, 207-210, 1991

A:Title: A new distinct group of 2 S albumins from rapeseed. Amino acid sequence of two

A:Reference number: S20350; MUID:92111741

A:Accession: S26636

A:Molecule type: protein

A:Residues: 1-106 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglutamic acid; seed; storage protein

F:1-31/Product: napin small chain #status experimental <SMA>

F:32-106/Product: napin large chain #status experimental <LAR>

F:32/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen

Query Match 44.8%; Score 82; DB 2; Length 106;
Best Local Similarity 40.5%; Pred. No. 0.00065;

Matches 15; Conservative 10; Mismatches 4; Indels 8; Gaps 1;

QY 4 QGCRGQIQEQNLRCQCEYIKQVSG-----QGP 32
Db 3 QGCRGQIQEQNLRCQCEYIKQVSG-----QGP 39

RESULT 3
S20350

napin n1a - rape

C:Species: Brassica napus (rape)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 04-Oct-1996

C:Accession: S20350; S20351

R:Monsalve, R.I.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R.

FEBS Lett. 295, 207-210, 1991

A:Title: A new distinct group of 2 S albumins from rapeseed. Amino acid sequence of two

A:Reference number: S20350; MUID:92111741

A:Accession: S20350

A:Molecule type: protein

A:Residues: 1-110 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglutamic acid; seed; storage protein

F:1-31/Product: napin small chain #status experimental <SMA>

F:32-110/Product: napin large chain #status experimental <LAR>

F:32/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen

Query Match 43.7%; Score 80; DB 2; Length 110;
Best Local Similarity 41.2%; Pred. No. 0.0013;
Matches 14; Conservative 11; Mismatches 5; Indels 4; Gaps 1;

QY 4 QGCRGQIQEQNLRCQCEYIKQVSG-----QGP 33

Db 3 QGCRGQIQEQNLRCQCEYIKQVSGGPRR 36

RESULT 4

S14947

2S albumin - Brazil nut

C:Species: Bertholletia excelsa (Brazil nut)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000

C:Accession: S14947

R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gr

Plant Mol. Biol. 16, 437-448, 1991

A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin

A:Reference number: S14946; MUID:91370890

A:Accession: S14947

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-154 <GAN>

A:Cross-references: EMBL:X54491; NID:917712; PIDN:CA38363.1; PID:917713

C:Genetics:

A:Introns: 67/3

C:Superfamily: wheat alpha-amylase inhibitor

Query Match 41.8%; Score 76.5; DB 2; Length 154;
Best Local Similarity 42.9%; Pred. No. 0.0052;
Matches 15; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

QY 2 SQGCRGQIQEQNLRCQCEYIKQVSG---QGP 33

Db 43 SQGCRGQIQEQNLRCQCEYIKQVSG---QGP 77

RESULT 5

T09878

albumin 2S storage protein precursor Mat5-A - upland cotton

C:Species: Gossypium hirsutum (upland cotton)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 29-Oct-1999

C:Accession: T09878

R:Galau, G.A.; Wang, H.Y.C.; Hughes, D.W.

submitted to the EMBL Data Library, January 1992

A:Description: Cotton Mat5-A (C164) gene and Mat5-D cDNAs encoding methionine-rich 2S

A:Reference number: Z16893

A:Accession: T09878

A>Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-139 <GAL>

A:Cross-references: EMBL:M66213; NID:9167358; PID:9167359

C:Genetics:

A:Gene: Mat5-A

C:Keywords: storage protein

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-139/Product: albumin 2S storage protein Mat5-A #status predicted <MAN>

Query Match 40.4%; Score 74; DB 2; Length 139;
Best Local Similarity 33.3%; Pred. No. 0.01;
Matches 10; Conservative 13; Mismatches 7; Indels 0; Gaps 0;

QY 2 SQGCRGQIQEQNLRCQCEYIKQVSGG 31

Db 32 NRDCSEQIRKQAHKHCQKYMEEELGEG 61

RESULT 6

B59346

seed storage protein Lec2SA1 small chain [imported] - tomato (fragment)

N:Alternate names: 2S albumin small subunit

C:Species: Lycopersicon esculentum (tomato)

C:Date: 31-Dec-2001 #sequence_revision 31-Dec-2001 #text_change 31-Dec-2001

C:Accession: B59346

R:Oguri, S.

submitted to the Protein Sequence Database, September 2000

A:Reference number: A59346
A:Accession: B59346
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-32 <OGU>
A:Experimental source: cv. cherry; seed
C:Complex: heterodimer of large (see PIR:A59346) and small chains, disulfide linked
C:Keywords: heterodimer; seed; storage protein

Query Match 39.8%; Score 71; DB 2; Length 32;
Best Local Similarity 42.3%; Pred. No. 0.0031;
Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 1 PSQGGCGQIQEQNLRCQCEYIKQV 26
DB 1 PQTSCQGGPQQAQQLRSCQFLKRR 26

RESULT 7

A25802
2S seed storage protein small chain - Brazil nut

C:Species: Bertholletia excelsa (Brazil nut)

C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 19-May-2000

C:Accession: A25802

R:Ampe, C.; Van Damme, J.; de Castro, L.A.B.; Sampalo, M.J.A.M.; Van Montagu, M.; Vandek

Eur. J. Biochem. 159, 597-604, 1986

A:Title: The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nu

A:Reference number: A91173; MUID:87004679

A:Accession: A25802

A:Molecule type: protein

A:Residues: 1-28 <AMP>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglutamic acid
F1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 38.8%; Score 71; DB 2; Length 28;
Best Local Similarity 52.0%; Pred. No. 0.005;
Matches 13; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 3 QQCGRQIQEQNLRCQCEYIKQV 27
DB 1 QOOCRFQMQRQQLSHCRMYMRQM 25

RESULT 8

S35591
major allergen small chain - leaf mustard

C:Species: Brassica juncea (leaf mustard)

C:Date: 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change 07-May-1999

C:Accession: S35591

R:Monsalve, R.I.; Gonzalez de la Pena, M.A.; Menendez-Arias, L.; Lopez-Otin, C.; Villal

Biochem. J. 293, 625-632, 1993

A:Title: Characterization of a new oriental-mustard (Brassica juncea) allergen, Bra j IF

A:Reference number: S35591; MUID:93356721

A:Accession: S35591

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-37 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

Query Match 38.8%; Score 71; DB 2; Length 37;
Best Local Similarity 44.8%; Pred. No. 0.0066;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQGP 32
DB 9 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 37

RESULT 9

S01791

allergen Sin a I small chain - white mustard

C:Species: Sinapis alba (white mustard)

C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 25-Oct-1996

C:Accession: S01791

R:Menendez-Arias, L.; Moneo, I.; Dominguez, J.; Rodriguez, R.

Eur. J. Biochem. 177, 159-166, 1988

A:Title: Primary structure of the major allergen of yellow mustard (Sinapis alba L.)

A:Reference number: S01791; MUID:89030681

A:Accession: S01791

A:Molecule type: protein

A:Residues: 1-39 <MEN>

A:Note: 6-Gly was also found

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed

Query Match 38.8%; Score 71; DB 2; Length 39;
Best Local Similarity 44.8%; Pred. No. 0.007;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQGP 32
DB 10 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 38

RESULT 10

S65447

allergen Sin a I - white mustard

N:Alternate names: allergen Sin a I large chain

C:Species: Sinapis alba (white mustard)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999

C:Accession: S65447; S01792

R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.

Eur. J. Biochem. 237, 827-832, 1996

A:Title: Expression in Escherichia coli of Sin a I, the major allergen from mustard.

A:Reference number: S65447; MUID:96235251

A:Accession: S65447

A:Molecule type: protein

A:Residues: 1-127 <GON>

A:Experimental source: seed

R:Menendez-Arias, L.; Moneo, I.; Dominguez, J.; Rodriguez, R.

Eur. J. Biochem. 177, 159-166, 1988

A:Title: Primary structure of the major allergen of yellow mustard (Sinapis alba L.)

A:Reference number: S01791; MUID:89030681

A:Accession: S01792

A:Molecule type: protein

A:Residues: 40-127 <MEN>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 127;
Best Local Similarity 44.8%; Pred. No. 0.023;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQNLRCQCEYIKQV--SGQGP 32
DB 10 CRKEFQQAQHLRACQQLHKKQAMQSGSGP 38

RESULT 11

PC1246

Sin a I allergen I precursor - white mustard (fragment)

C:Species: Sinapis alba (white mustard)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999

C:Accession: PC1246

R:de la Pena, M.A.G.; Villalba, M.; Garcia-Lopez, J.L.; Rodriguez, R.

Biochem. Biophys. Res. Commun. 190, 648-653, 1993

A:Title: Cloning and expression of the major allergen from yellow mustard seeds, Sin

A:Reference number: PC1246; MUID:93151856

A:Accession: PC1246

A:Molecule type: DNA

A:Residues: 1-145 <DLP>
A:Experimental source: seed
A>Note: the authors translated the codon GAG for residue 45 as Gln
A>Note: the species is designated as "yellow mustard"
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-54/Domain: propeptide #status predicted <PRO>
F:55-145/Product: sin a I allergen 1 (fragment) #status predicted <MAT>

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFOQAQHLRACQOWLHKQAMQSGSGP 38

RESULT 12

565478
allergen Sin a I (clone SIN5) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65478
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65478
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91798; NID:g1009441; PIDN:CAA62908.1; PID:g1009442
A:Experimental source: seed
C:Genetics:
A:Gene: sin5
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFOQAQHLRACQOWLHKQAMQSGSGP 38

RESULT 13

565481
allergen Sin a I (clone SIN3) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65481
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65481
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91801; NID:g1009437; PIDN:CAA62911.1; PID:g1009438
A:Experimental source: seed
C:Genetics:
A:Gene: sin3
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFOQAQHLRACQOWLHKQAMQSGSGP 38

RESULT 14

565482
allergen Sin a I (clone SIN4) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65482
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65482
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91802; NID:g1009439; PIDN:CAA62912.1; PID:g1009440
A:Experimental source: seed
C:Genetics:
A:Gene: sin4
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFOQAQHLRACQOWLHKQAMQSGSGP 38

RESULT 15

565480
allergen Sin a I (clone SIN2) - white mustard (fragment)
C:Species: Sinapis alba (white mustard)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S65480
R:Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.
Eur. J. Biochem. 237, 827-832, 1996
A:Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.
A:Reference number: S65447; MUID:96235251
A:Accession: S65480
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-145 <GON>
A:Cross-references: EMBL:X91800; NID:g1009435; PIDN:CAA62910.1; PID:g1009436
A:Experimental source: seed
C:Genetics:
A:Gene: sin2
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed; storage protein

Query Match 38.8%; Score 71; DB 2; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.027;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFOQAQHLRACQOWLHKQAMQSGSGP 38

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
57.312 Million cell updates/sec

Title: US-09-913-351-4
Perfect score: 183
Sequence: 1 PSQCGCRGQIQEQONLRQCEYIKQGVSGGPRR 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	258	1 2SS_RICCO	P01089 ricinus com
2	80	43.7	110	1 2SS1_BRANA	P24565 brassica na
3	72	39.3	129	1 ALL1_BRAJU	P80207 brassica ju
4	71	38.8	127	1 ALL1_SINLU	P15322 sinapis alu
5	71	38.8	130	1 ITRY_SINAR	P38057 sinapis arv
6	70	38.3	146	1 2SS1_BRANA	P04403 bertholleti
7	69	37.7	104	1 2SS1_BRANA	P80351 caparis ma
8	69	37.7	178	1 2SS2_BRANA	P01090 brassica na
9	68	37.2	178	1 2SS2_BRANA	P27740 brassica na
10	68	37.2	180	1 2SS3_BRANA	P17333 brassica na
11	67.5	36.9	37	1 CG25_LUPAN	P09930 lupinus ang
12	67	36.6	104	1 2SS3_CAPPA	P80352 caparis ma
13	66.5	36.3	100	1 2SS4_CAPPA	P80353 caparis ma
14	66	36.1	133	1 2SS1_BRANA	P01091 brassica na
15	66	36.1	164	1 2SS1_BRANA	P15457 arabidopsis
16	66	36.1	186	1 2SS1_BRANA	P09893 arabidopsis
17	64.5	35.2	166	1 2SS4_ARATH	P15460 arabidopsis
18	63.5	34.7	164	1 2SS3_ARATH	P15458 arabidopsis
19	62.5	34.2	170	1 2SS2_ARATH	P15458 arabidopsis
20	58.5	32.0	158	1 2SS1_BRANA	P19594 glycine max
21	57	31.1	182	1 INB_MOUSE	P01575 mus musculu
22	56.5	30.9	705	1 FXPI_MOUSE	P58466 mus musculu
23	56	30.6	525	1 VCL_THREC	O43358 theobroma c
24	55.5	30.3	304	1 GDB1_WHEAT	P04729 triticum ae
25	55	30.1	904	1 AAG4_CHICK	Q50734 gallus gall
26	54	29.5	313	1 GAT7_WHEAT	P04727 triticum ae
27	53.5	29.2	295	1 GLTQ_WHEAT	P16335 triticum ae
28	53	29.0	144	1 IAAE_HORVU	P01086 hordeum vul
29	53	29.0	4349	1 DYHC_FUSOO	P78716 fusarium so
30	52.5	28.7	1060	1 ZFHL_DROME	P81816 drosophila
31	52	28.4	155	1 2SS2_CAPPA	P20233 caparis ma
32	52	28.4	290	1 RAS1_CANAL	Q94890 drosophila
33	51	27.9	482	1 NF31_NABFO	P42661 naegleria f

34	51	27.9	905	1 SNF5_YEAST	P18480 saccharomyc
35	51	27.9	2038	1 FSH_DROME	P13709 drosophila
36	51	27.9	3828	1 TRX_DROVI	O24742 drosophila
37	50.5	27.6	289	1 HOG3_HORVU	P80198 hordeum vul
38	50.5	27.6	295	1 2SS5_HELAN	P15461 helianthus
39	50.5	27.6	648	1 GLTQ_WHEAT	P10387 triticum ae
40	50.5	27.6	660	1 GLT3_WHEAT	P08488 triticum ae
41	50.5	27.6	1477	1 HRT7_HYDRA	O25197 hydra atten
42	50.5	27.6	3703	1 ABFL_HUMAN	O15911 homo sapien
43	50	27.3	435	1 NPL6_YEAST	P32832 saccharomyc
44	50	27.3	544	1 H15_DROME	Q94890 drosophila
45	50	27.3	873	1 RX_DROME	Q9w2q1 drosophila

ALIGNMENTS

```

RESULT 1
ID 2SS_RICCO STANDARD: PRT: 258 AA.
AC P01089;1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm.
RX MEDLINE=91108729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein
RT may be processed into two different heterodimeric storage proteins.";
RL Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
RT gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
RT protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6615448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
RT inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY 2 DISULFIDE BONDS.
CC -1- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CC CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/

```

```
CC or send an email to license@lsb.sib.ch).
-----
DR EMBL: X54158; CAA38097.1; .
DR PIR: A01328; RZCS.
DR PIR: S11499; S11499.
DR PIR: S11500; S11500.
DR PIR: S11501; S11501.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_tryp_aml1_inh.
DR Pfam: PF00234; tryp_alpha_aml1_2.
DR SMART: SM00499; AAI: 2.
KW Seed storage protein; Signal; Allergen.
FT SIGNAL 1 21
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT PROPEP 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD_RES 194 194 PYROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 MISSING (IN REF. 3).
FT CONFLICT 234 234 D -> N (IN REF. 3).
FT CONFLICT 255 255 E -> Q (IN REF. 3).
SQ SEQUENCE 258 AA; 29290 MW; 2787ACFC50E41072 CRC64;

Query Match 100.0%; Score 183; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 8.4e-17;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PSQGCRGQIQEONLRCCQCEYIKQVSGGQPRR 34
Db 157 PSQGCRGQIQEONLRCCQCEYIKQVSGGQPRR 190

RESULT 2
2SST_BRANA
ID 2SST_BRANA STANDARD; PRT; 110 AA.
AC P24565;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Napin IA and IB small chain and large chains.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=92111741; Pubmed=1765156;
RA Monsalve R.I., Lopez-Olin C., Villalba M., Rodriguez R.;
RT "A new distinct group of 2 S albumins from rapeseed. Amino acid
sequence of two low molecular weight naplins."
RL FEBS Lett. 295:207-210(1991).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
ITS MATURATION
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY DISULFIDE BONDS.
CC -1- LINKED BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: NAPIN IA AND IB ARE MINOR COMPONENT OF SEED 2S
ALBUMIN.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIN.
DR Prodom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein.
FT CHAIN 1 31 SMALL CHAIN.
FT CHAIN 31
```

```
FT NON_CONS 31 32
FT CHAIN 32 110 LARGE CHAIN.
FT VARIANT 31 31 MISSING (IN MINOR FORM "SM", LESS
THAN 7%).
FT VARIANT 37 37 E -> Q (IN NAPIN IB).
FT VARIANT 107 110 MISSING (IN NAPIN IB).
SQ SEQUENCE 110 AA; 12691 MW; 3A2938ADA2C1E995 CRC64;

Query Match 43.7%; Score 80; DB 1; Length 110;
Best Local Similarity 41.2%; Pred. No. 0.00081;
Matches 14; Conservative 11; Mismatches 5; Indels 4; Gaps 1;

QY 4 QGCRGQIQEONLRCCQCEYIKQVSG---QGPR 33
Db 3 QKCRFQOEQHLRACQOWLRQAMQSGSPQ 36

RESULT 3
ALL1_BRADU
ID ALL1_BRADU STANDARD; PRT; 127 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Allergen Bra j 1-E, small and large chains (Bra j 1).
OS Brassica juncea (Leaf mustard) (Indian mustard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3707;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=93356721; Pubmed=7688955;
RA Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Arias L.,
RA Lopez-Olin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope."
RL Biochem. J. 293:625-632(1993).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY TWO DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NAPIN.
DR Prodom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Allergen; Seed storage protein.
FT CHAIN 1 37 SMALL CHAIN.
FT NON_CONS 37 38
FT CHAIN 38 129 LARGE CHAIN.
FT VARIANT 6 6 F -> I.
FT VARIANT 20 20 R -> K.
SQ SEQUENCE 129 AA; 14644 MW; D6F28E03F62B08F8 CRC64;

Query Match 39.3%; Score 72; DB 1; Length 129;
Best Local Similarity 43.3%; Pred. No. 0.01;
Matches 13; Conservative 8; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEONLRCCQCEYIKQV--SGGPR 33
Db 9 CRKEFQQAHLRACQOWLRQAMQSGSPQ 38

RESULT 4
ALL1_SINNL
ID ALL1_SINNL STANDARD; PRT; 127 AA.
AC P15322;
DT 01-APR-1990 (Rel. 14, Created)
```

DT	01-APR-1990 (Rel. 37, last sequence update)
DT	15-DEC-1998 (Rel. 37, last annotation update)
DE	Allergen Sin a 1, small and large chains (sin a I).
OS	Sinapis alba (White mustard) (Brassicaceae).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosida
OX	eumustoids II; Brassicales; Brassicaceae; Sinapis.
RN	NCBI_TaxID=3728;
RP	[1]
SEQUENCE	
RC	T1SDP-Seed:
RX	MEDLINE=89030681; PubMed=3181153;
RA	Menendez-Arias L., Moneo I., Dominguez J., Rodriguez R.;
RT	"Primary structure of the major allergen of yellow mustard (Sinapis
RL	alba L.) seed, Sin a I.";
RL	Eur. J. Biochem. 177:159-166(1988).
CC	-1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC	-1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC	BONDS.
CC	-1- DISEASE: ALLERGIC DISEASE, CABBAGE ALLERGY.
CC	-1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR	PIR: S01792; S01792.
DR	PIR: S01791; S01791.
DR	Interpro: IPR0003612; AAT.
DR	Interpro: IPR000617; Napin.
DR	Pfam: PF01631; Seedscore_2s; 1.
DR	PRINTS: PR00496; Napin.
DR	PRODOM: PD002458; Napin; 1.
DR	SMART: SM00499; AAT; 1.
KW	Allergen; Seed storage protein.
FT	CHAIN 1 39
FT	NON CONS 39 40
FT	CHAIN 40 127
FT	CHAIN 40 127
FT	VARIANT 6 6
FT	LARGE CHAIN.
FT	CHAIN 127 AA; 14180 MW; 4CD920284F04EED CRC64;
QO	SEQUENCE

Query Match Similarity 38.8%; Score 71; DB 1; Length 127;
 Best Local Similarity 44.8%; Pred. NO. 0.014;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1.

QY 6 CRRGQIQEQQNLRCQCEYKQGV--SGQCP 32
 || : | : | | | : : : | | | |
 Db 10 CRKEFQQAQHLRACQGWLHKQAMQSGSCP 38

RESULT 5
 ITTRY_SINAR
 ID ITTRY_SINAR STANDARD; PRT; 130 AA.

AC P38057;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trypsin Inhibitor (TISA).
 OS Sinapis arvensis (Charlock).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC Spermatophyta: Magnoliophyta: eudicotyledons, core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Sinapis.
 OX NCBI_TaxID=25728;
 RN [1]
 RM
 RP
 RC TISSUE=Seed;
 RX MEDLINE=94350545; PubMed=8070965;
 RA Svendsen I.B., Micolova D., Goshay I., Genov N.;
 RT "Primary structure, spectroscopic and inhibitory properties of a two-
 chain trypsin inhibitor from the seeds of charlock (*Sinapis arvensis*
 L.), a member of the napin protein family.";
 RL Int. J. Pept. Protein Res. 43:425-430(1994).
 CC -1 FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10(-6) M.
 CC -1 SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
 CC BONDS.
 CC -1 SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC InterPro: IPR003612; AAI.

Query Match	38.8%	Score 71:	DB 1:	Length 130:
Best Local Similarity	44.8%	Pred. NO. 0.014:		
Matches 13: Conservative	7:	Mismatches	7:	Indels 2: Gaps 1
DR	InterPro: IPR000617; Napin.			
DR	Pfam; PF01631; Seedstore_25; 1.			
DR	PRINTS; PR00496; NAPIN.			
DR	ProDom; PD002498; Napin; 1.			
DR	SMART; SM00499; AAI; 1.			
KW	Seed storage protein; Multigene family; Polymorphism.			
FT	CHAIN	1	39	
FT	NON-CONS	39	40	
FT	CHAIN	40	130	
FT	VARIANT	32	32	LARGE CHAIN.
FT	VARIANT	53	57	R -> M.
FT	VARIANT	73	73	MISSING (IN FORM II).
FT	VARIANT	77	77	A -> S.
FT	VARIANT	81	81	K -> R.
FT	VARIANT	87	87	Q -> R.
FT	VARIANT	89	89	O -> Q.
FT	VARIANT	91	91	G -> Q.
FT	VARIANT	97	97	E -> M.
FT	VARIANT	98	98	I -> V.
FT	VARIANT	99	99	R -> S.
FT	VARIANT	106	106	T -> K.
FT	VARIANT	123	123	N -> Q.
FT	VARIANT	124	124	K -> G.
FT	VARIANT	126	126	M -> V.
SEQ	SEQUENCE	130 AA;	14682 MW;	EC02E4B26D180DF2 CRC64;

```

QY      6  CREGIOEQONILROCEQYIKQV--SQGCP 32
      11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DQ      10 CRKEFQQAQHLRACQQLWLRQARQSSGSP 38
      10 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT      6
ID      2SS_BEREX      STANDARD;      PRT;      146 AA.
AC      P04403;
DT      20-MAR-1987 (Rel. 04, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      2S sulfur-rich seed storage protein precursor (Allergen Ber e 1).
GN      BE2S1 AND BE2S2.
OS      Bertholletia excelsa (Brazil nut).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; Ericales; Lecythidaceae; Bertholletia.
OX      NCBI_TaxID=3645;
LN      1
RA      Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M.;
RT      "Cloning and sequence analysis of a cDNA encoding a Brazil nut protein
RT      exceptionally rich in methionine."
RL      Plant Mol. Biol. 8:239-250(1987).
LN      2
RA      SEQUENCE FROM N.A.
RP      Bassuener R.;
RL      Submitted (DEC-1990) to the EMBL/Genbank/DBJ databases.
LN      3
RA      SEQUENCE FROM N.A.
RP      MEDLINE=91370890; PubMed=1840683;
RA      Gander E.S., Holmstrom K.O., de Paiva G.R., de Castro L.A.B.,
RA      Carneiro M., Grossi de Sa M.F.;
RT      "Isolation, characterization and expression of a gene coding for a 2S
RT      albumin from Bertholletia excelsa (Brazil nut).";
RL      Plant Mol. Biol. 16:437-448(1991).
LN      4
RA      SEQUENCE OF 37-64 AND 70-142.
RP      MEDLINE=87004679; PubMed=3758080;
RA      Ampe G., van Damme J., de Castro L.A.B., Sampao M.J.A.M.,
RA      van Montagu M., Vandekerckhove J.;

```

```

RT      "The amino-acid sequence of the 2S sulphur-rich proteins from seeds
FT      of Brazil nut (Bertholletia excelsa H.B.K.).";
RL      Eur. J. Biochem. 159:597-604(1986).
CC      -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC      -1- SUBUNIT: THE MAJURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC      LINKED BY DISULFIDE BONDS.
CC      -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
-----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
DR      EMBL; M17146; AAA3010.1; -.
DR      EMBL; X57027; CAA40343.1; -.
DR      EMBL; X57028; CAA40344.1; -.
DR      EMBL; X54490; CAA38362.1; -.
DR      EMBL; X54491; CAA38363.1; ALT-SEQ.
DR      EMBL; A13818; CAA01131.1; -.
DR      PIR; S06252; S06252.
DR      PIR; A25802; A25802.
DR      PIR; B25802; B25802.
DR      PIR; S14946; S14946.
DR      PIR; S14479; S14479.
DR      PIR; S21640; S21640.
DR      InterPro: IPR003612; AAI.
DR      InterPro: IPR001768; Cereals_tryp_amyl_1nh.
DR      Pfam; PF00234; tryp_alpha_amyl; 1.
DR      SMART; SM00499; AAT; 1.
KW      Seed storage protein; Signal; Allergen.
FT      SIGNAL          1        22
FT      PROPEP         23        36
FT      CHAIN          37        64       SMALL CHAIN.
FT      PROPEP         65        69
FT      CHAIN          70       142       LARGE CHAIN 1B.
FT      PROPEP        143       146
FT      MOD_RES        37        37
FT      VARIANT        91        91       PYRROLIDONE CARBOXYLIC ACID.
FT      CONFLICT       38        39       S->E (IN VARIANT 1A).
FT      CONFLICT       91        91       EE->QQ (IN REF. 4).
FT      CONFLICT       122       122       L->M (IN REF. 4).
FT      CONFLICT       126       126       I->L (IN REF. 4).
SQ      SEQUENCE      146 AA; 16911 MM; A7DF778FD766410D CRC64;
-----
Query Match              38.3%; Score 70; DB 1; Length 146;
Best Local Similarity    40.5%; Pred. NO. 0.022;
Matches 15; Conservative 7; Mismatches 11; Indels 4; Gaps 1;
QY      2 SQQGCGGQIOEQNLROCCQEYIKQGVSGG---RRR 34
DB      36 NQEBCEQMQRQMLSHCRMYNRQGMESPPTMR 72
-----
RESULT 7
AC      2SS1_CAPMA STANDARD; PRT; 104 AA.
ID      2SS1_CAPMA
PD      P80351;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Mabinlin I-L, A and B chains (MAB I) (Sweet protein).
OS      Capparis maseikai (Mabiniang).
OC      Eudaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosoids II; Brassicales; Brassicaceae; Caparis.
OX      NCBI_TaxId=13395;
RN      [1]
RP      SEQUENCE.
RC      TITSDU-Seed;
RX      MEDLINE=94333405; PubMed=8055976;
```

RA Nirasawa S., Nishino T., Katsuhira M., Uesugi S., Hu Z., Kurihara Y.,
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue."
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING
CC ACTIVITY. THIS FORM IS NOT HEAT STABLE.
CC -1- SUBUNIT: HETEROOLIGOMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; NApin.
DR InterPro: IPR000617; NApin.
DR Pfam: PF01631; Seedstore_2s; 1.
DR PRINTS: PR00496; NApin.
DR ProDom: PD002498; NApin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 32 MABINLIN I, A CHAIN.
FT FT NON CONS 32 33
FT CHAIN 33 104 MABINLIN I, B CHAIN.
FT DISULFID 4 53
FT DISULFID 17 42
FT DISULFID 43 91
FT DISULFID 55 99
SO SEQUENCE 104 AA; 12284 MW; 15644A4B50D5EE60 CRC64;

QY 6 CRGQIQEQNLRQCCQEVYIKQVSCGG 31
||| : ||| ||| : :
Db 4 CRQFOQHQLRACQRYIRRAQRCG 29

RESULT 8
2SS2_BRANA STANDARD; PRT; 178 AA.
ID 2SS2_BRANA
AC P01090;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE NApin 2 precursor (1.7S seed storage protein).
OS Brassica napus (Rapeseed).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC eunotids II: Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308224; PubMed=3624251;
RA Josefson L.-G., Lenman M., Ericson M.L., Rask L.;
RT "Structure of a gene encoding the 1.7 S storage protein, napin, from
RT Brassica napus."
RL J. Biol. Chem. 262:12196-12201(1987).
RN [2]
RP REVISIONS.
RA Josefson L.-G.;
RL Submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87033655; PubMed=3771543;
RA Ericson M.L., Reedin J., Lenman M., Glimelius K., Josefson L.-G.,
RA Rask L.;
RT "Structure of the rapeseed 1.7 S storage protein, napin, and its
RT precursor."
RL J. Biol. Chem. 261:14576-14581(1986).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TOWER;
RX MEDLINE=84113267; PubMed=6689334;
RA Crouch M.L., Tenbarger K.M., Simon A.E., Ferl R.;

"cDNA clones for Brassica napus seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide.";
 J. Mol. Appl. Genet. 2:273-283(1983).
 - FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING ITS MATURATION.
 - SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN LINKED BY DISULFIDE BONDS.
 - TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 - SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; K01545; AAA33006.1; -
 CC EMBL; J02586; AAA3297.1; -
 CC EMBL; J02798; AAA87348.1; -
 CC PIR; A01329; NMRP2.
 CC PIR; A29801; A29801.
 CC PIR; A25997; A25997.
 CC InterPro; IPR003612; AAI.
 CC InterPro; IPR001768; Cereal-trypan-amy1_inh.
 CC InterPro; IPR000617; Napin.
 CC Pfam; PF00234; trypan-alpha-amy1; 1.
 CC PRINTS; PR00496; Napin.
 CC ProDom; PD002498; Napin; 1.
 CC SMART; SM00499; AAI; 1.
 CC Seed storage protein; Signal; Multigene family.
 CC SIGNAL 1 21
 CC PROPEP 22 38
 CC CHAIN 39 74 SMALL CHAIN.
 CC PROPEP 75 94
 CC CHAIN 95 175 LARGE CHAIN.
 CC CONFLICT 37 37 D -> N (IN REF. 4).
 CC CONFLICT 76 76 S -> N (IN REF. 4).
 CC SEQUENCE 178 AA; 20104 MW; 734E561971B539FF CRC64;
 SQ
 Query Match 37.7%; Score 69; DB 1; Length 178;
 Best Local Similarity 44.8%; Pred. No. 0.035;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;
 OY 6 CRGQIOEOONLRQCOEYIKQV--SGGPG 32
 ID 2SS3_BRANA STANDARD; PRT; 178 AA.
 AC P27740.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Napin B precursor (1.7S seed storage protein).
 GN NABP.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 CC NCB1_TaxID=3708;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. SVALOFS KARAT 20516-K;
 RA MEDLINE=91231016; PubMed=2029903;
 RA Ericson M.L., Muren E., Gustavsson H.O., Josefsson L.G., Rask L.;
 "Analysis of the promoter region of napin genes from Brassica napus

demonstrates binding of nuclear protein in vitro to a conserved sequence motif.";
 Eur. J. Biochem. 197:741-746(1991).
 - FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING ITS MATURATION.
 - SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN LINKED BY DISULFIDE BONDS.
 - TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 - SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; X58142; CAA41150.1; -
 CC PIR; S15382; S15382.
 CC InterPro; IPR003612; AAI.
 CC InterPro; IPR001768; Cereal-trypan-amy1_inh.
 CC InterPro; IPR000617; Napin.
 CC Pfam; PF00234; trypan-alpha-amy1; 1.
 CC PRINTS; PR00496; Napin.
 CC ProDom; PD002498; Napin; 1.
 CC SMART; SM00499; AAI; 1.
 CC Seed storage protein; Signal; Multigene family; Embryo.
 CC SIGNAL 1 21
 CC PROPEP 22 38 BY SIMILARITY.
 CC CHAIN 39 74 BY SIMILARITY.
 CC PROPEP 75 94 SMALL CHAIN (BY SIMILARITY).
 CC CHAIN 95 178 LARGE CHAIN (BY SIMILARITY).
 CC SEQUENCE 178 AA; 20114 MW; 96CE0ADB7CD966E9 CRC64;
 SQ
 Query Match 37.2%; Score 68; DB 1; Length 178;
 Best Local Similarity 41.4%; Pred. No. 0.048;
 Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 1;
 OY 6 CRGQIOEOONLRQCOEYIKQV--SGGPG 32
 ID 2SS3_BRANA STANDARD; PRT; 180 AA.
 AC P17333.
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Napin precursor (1.7S seed storage protein).
 GN NABP.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 CC NCB1_TaxID=3708;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CV. WESTAR; TISSUE=Leaf;
 RA MEDLINE=91346654; PubMed=2102844;
 RA Baszczyński C.L., Fallis L.;
 "Isolation and nucleotide sequence of a genomic clone encoding a new Brassica napus napin gene.";
 Plant Mol. Biol. 14:633-635(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. TOPAS;
 RA MEDLINE=95161697; PubMed=7858212;
 RA Boutilier K.A., Gines M.J., Demoor J.M., Huang B.,

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgatterer M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnselt J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Gymonprez B., Chuang Y.-J., Vandebussche F.,
 RA Breken M., Welfjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzengger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holtzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
 RA Benesler S., Hempel S., Feldpausch M., Lambert S., Van den Daele H.,
 RA De Keyser A., Busschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Petrett A., Rajandream M.A., Lyne M., Benes V., Neumann S.,
 RA Botkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Ungley F., Clabaud G., Mendenlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cherdorf T., Weber N., Vandepol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Frishman D., Haase D., Lemcke K., Meyers H.-W., Stocker S.,
 RA Zaccaria P., Bervan M., Wilson R.K., de la Bastide W., Habermann K.,
 RA Parnell L., Dedha N., Gnoj L., Schütz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Marx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spleth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RA thaliana."
 RT Nature 402:769-777(1999).
 CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
 CC ARABIDOPSIS.
 CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M22032; AAA32743.1; -
 DR EMBL; Z24745; CAA80870.1; -
 DR EMBL; A13820; CAA01132.1; -
 DR EMBL; A1035680; CAB38844.1; -
 DR EMBL; A161566; CAB79569.1; -
 DR PIR; JA0161; NMMU1.
 DR PIR; S34676; S34676.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal_1tryp_aml_1nh.
 DR InterPro: IPR000617; Napin.
 DR Pfam; PF00234; tryp_alpha_aml1; 1.
 DR PRINTS; PR00496; NAPIN.

DR ProDom: PD002498; Napin; 1.
 DR SMART; SM00499; AAI; 1.
 KW Seed storage protein; Albumin; Signal; Multigene family.
 FT SIGNAL 1 21
 FT PROPEP 22 37
 FT CHAIN 38 73 2S SEED STORAGE PROTEIN 1 SMALL SUBUNIT.
 FT PROPEP 74 83
 FT CHAIN 84 162 2S SEED STORAGE PROTEIN 1 LARGE SUBUNIT.
 FT PROPEP 163 164
 FT PROPEP 164 164
 SQ SEQUENCE 164 AA; 19014 MW; 2BF28CB474D9832B CRC64;

Query Match 36.1%; Score 66; DB 1; Length 164;
 Best Local Similarity 42.3%; Pred. No. 0.08;
 Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 PSQCRCRGQIOEQONTROCGEYIKQ 26
 DB 41 PMRCRCRFQKQKHTRACQQLMQ 66

Search completed: September 9, 2002, 12:46:06
 Job time: 308 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:37 : Search time 77.42 Seconds
(without alignments)
75.973 Million cell updates/sec

Title: US-09-913-351-4
Sequence: 1 PS00GCRGQIOEQONLRQCCQEVTKQVSGGQPRR 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp_unclassified: *
15: sp_rv1rus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	183	100.0	35	10	Q9S873	Q9S873 ricinus com
2	88	48.1	139	10	P93198	P93198 Juglans reg
3	74	40.4	139	10	Q39795	Q39795 gossypium h
4	72	39.3	125	10	P80208	P80208 brassica na
5	71	38.8	40	10	Q9S9F2	Q9S9F2 brassica na
6	71	38.8	145	10	Q41196	Q41196 sinapis alb
7	71	38.8	145	10	Q41281	Q41281 sinapis alb
8	71	38.8	145	10	Q41277	Q41277 sinapis alb
9	71	38.8	145	10	Q41278	Q41278 sinapis alb
10	71	38.8	145	10	Q41279	Q41279 sinapis alb
11	71	38.8	145	10	Q41280	Q41280 sinapis alb
12	71	38.8	152	10	Q41168	Q41168 rapanus sa
13	71	38.8	174	10	Q41169	Q41169 rapanus sa
14	71	38.8	180	10	Q42469	Q42469 brassica na
15	70	38.3	41	10	Q9S9F1	Q9S9F1 brassica na
16	70	38.3	139	10	Q39787	Q39787 gossypium h

17	70	38.3	146	10	Q9LR02	Q9LR02 bertholletii
18	69	37.7	178	10	Q42490	Q42490 brassica ol
19	69	37.7	178	10	Q39344	Q39344 brassica na
20	68	37.2	178	10	Q42473	Q42473 brassica ca
21	68	37.2	184	10	Q42444	Q42444 brassica ca
22	67	36.6	178	10	Q42491	Q42491 brassica na
23	66.5	36.3	153	10	Q99235	Q99235 lupinus ang
24	66	36.1	39	10	Q9S802	Q9S802 rapanus ang
25	66	36.1	39	10	Q9S802	Q9S802 rapanus ang
26	65	35.5	178	10	Q9S339	Q9S339 brassica ca
27	65	35.5	39	10	Q9S801	Q9S801 rapanus sa
28	63	34.4	39	10	Q9S803	Q9S803 rapanus sa
29	62	33.9	155	10	Q41167	Q41167 rapanus sa
30	61	33.3	178	10	Q42413	Q42413 brassica ju
31	61	33.3	323	10	Q39928	Q39928 helianthus
32	61	33.3	528	10	Q80838	Q80838 arabidopsis
33	61	33.3	528	10	Q94Bv4	Q94Bv4 arabidopsis
34	60	32.8	882	11	Q9D4H4	Q9D4H4 mus musculu
35	59.5	32.5	41	10	Q9S874	Q9S874 ricinus com
36	59	32.2	30	10	Q9S8Y8	Q9S8Y8 rapanus sa
37	58	31.7	295	10	Q946W1	Q946W1 zea mays (m
38	58	31.7	378	5	Q27383	Q27383 caenorhabdi
39	57.5	31.4	386	5	Q9W353	Q9W353 drosophila
40	57.5	31.4	618	5	Q9XVY8	Q9XVY8 caenorhabdi
41	57.5	31.4	635	5	Q18094	Q18094 caenorhabdi
42	57	31.1	388	5	Q16500	Q16500 caenorhabdi
43	57	31.1	388	5	Q16501	Q16501 caenorhabdi
44	57	31.1	438	5	Q16502	Q16502 caenorhabdi
45	56.5	30.9	445	5	Q16511	Q16511 caenorhabdi
			141	10	Q39649	Q39649 cucurbita s

ALIGNMENTS

RESULT 1
ID Q9S873 PRELIMINARY; PRT; 35 AA.
AC Q9S873;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 4 KDA NAPIIN-LIKE PROTEIN SMALL CHAIN (FRAGMENT).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxId=3988;
RN (1)
RP SEQUENCE.
RX MEDLINE-97135090; PubMed-8980648;
RA Neumann G.M., Condron R., Polya G.M.;
RT "Purification and sequencing of napin-like protein small and large
RT chains from Momordica charantia and Ricinus communis seeds and
RT determination of sites phosphorylated by plant Ca(2+)-dependent
RT protein kinase.";
RL Biochim. Biophys. Acta 1298:223-240(1996).
SQ SEQUENCE 35 AA; 4044 MW; 992B9BA8D6A68FD3 CRC64;

Query Match 100.0%; Score 183; DB 10; Length 35;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PS00GCRGQIOEQONLRQCCQEVTKQVSGGQPRR 34
DB 1 PS00GCRGQIOEQONLRQCCQEVTKQVSGGQPRR 34
RESULT 2
ID P93198 PRELIMINARY; PRT; 139 AA.
AC P93198;
DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALBUMIN SEED STORAGE PROTEIN (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND.
RA Teuber S.S., Dandekar A.M., Peterson W.R.;
RT Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; 066866; AAB41308.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1ryp_amy_1nh.
DR InterPro; IPR000480; Glutelin.
DR Pfam; PF00234; 1ryp_alpha_amy; 1.
DR PRINTS; PR00211; GLUTELIN.
DR SMART; SM00499; AAI; 1.
DR NON_TER 1
SQ SEQUENCE 139 AA; 16373 MW; 02D0E55E67164F23 CRC64;

Query Match	48.1%	Score 88	DB 10	Length 139
Best Local Similarity	57.1%	Pred. No. 6	4e-05	
Matches	16	Conservative	3	Mismatches 9; Indels 0; Gaps 0
QY	4	GGCGGQIDPQNLNRCCQETIKQVSGG	31	
		:		
Db	37	EGCGRIDQRLNLNHCQYLLRQSSRG	64	

[illegible]

Query Match	40.4%	Score 74	DB 10	Length 139
Best Local Similarity	33.3%	Pred. No. 0.0063		
Matches 10, Conservative 13		Mismatches 7	Indels 0	Gaps 0

QY 2 SQGCRGQIQEQNLRCQCEYIKQVSGG 31

Db 32 NRDSCEQDIRKQAH LKHCQKYMEELGEG 61

RESULT 4

AC P80208;

DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

NCBI_TaxID=3708;

RC TISSUE=SEED; MEDTUNE=07300373. PubMed=0340377.

RA Fernandez C., Villalba M., Rodriguez R.;

CLIN. EXP. ALLERGY 21:833-841(1991).

CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC ITS MATURATION.

CC -1- TISSUE SPECIFICITY: COLLEBONS AND THE AXIS.
CC -1- STIMULABILITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.

```
DR InterPro: IPR000617: Nucleic acid binding domain
DR InterPro: IPR001768; Cereals_c1yp_damyl_1_nucl
DR InterPro: IPR000617: Nucleic acid binding domain
```

DR	PR002498:	Näpin: 1.
DR	PR00450:	NAFLN:
PRINIS:		

nm seen storage protein; water-soluble; small chain.

Case	Age	Sex	Occupation	Duration of illness	Site of lesion	Pathological changes	Microscopic findings	Diagnosis
1	45	Male	Farmer	10 years	Brain	Chronic	Microscopic	Chronic
2	55	Female	Housewife	5 years	Brain	Chronic	Microscopic	Chronic
3	65	Male	Teacher	15 years	Brain	Chronic	Microscopic	Chronic
4	75	Female	Retired	20 years	Brain	Chronic	Microscopic	Chronic
5	85	Male	Farmer	25 years	Brain	Chronic	Microscopic	Chronic
6	95	Female	Housewife	30 years	Brain	Chronic	Microscopic	Chronic
7	105	Male	Farmer	35 years	Brain	Chronic	Microscopic	Chronic
8	115	Female	Housewife	40 years	Brain	Chronic	Microscopic	Chronic
9	125	Male	Farmer	45 years	Brain	Chronic	Microscopic	Chronic
10	135	Female	Housewife	50 years	Brain	Chronic	Microscopic	Chronic
11	145	Male	Farmer	55 years	Brain	Chronic	Microscopic	Chronic
12	155	Female	Housewife	60 years	Brain	Chronic	Microscopic	Chronic
13	165	Male	Farmer	65 years	Brain	Chronic	Microscopic	Chronic
14	175	Female	Housewife	70 years	Brain	Chronic	Microscopic	Chronic
15	185	Male	Farmer	75 years	Brain	Chronic	Microscopic	Chronic
16	195	Female	Housewife	80 years	Brain	Chronic	Microscopic	Chronic
17	205	Male	Farmer	85 years	Brain	Chronic	Microscopic	Chronic
18	215	Female	Housewife	90 years	Brain	Chronic	Microscopic	Chronic
19	225	Male	Farmer	95 years	Brain	Chronic	Microscopic	Chronic
20	235	Female	Housewife	100 years	Brain	Chronic	Microscopic	Chronic

Best Local Similarity 43.38; Pred. No. 0.011;

6 CRGQIQEQNLRCQEQYIKQV--SGQGP 33

Q9S9F2

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DE NAPIN SHORT CHAIN S3A=CALMODULIN ANTAGONIST/CALCIUM-DEPENDENT PROTEIN

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Equisetum

OX NCBI_TaxID=3708;
 OX f11

RX MEDLINE=96283790; PubMed=8679670;
Newman C M, Gordon B, Thompson

RT napin small chains that are calmodulin antagonists and substrates for
RT plant calcium-dependent protein kinase."
RL Biochim. Biophys. Acta 1295:23-33(1996).
SQ SEQUENCE 40 AA; 4620 MW; BC7E47D33800B593 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.0045;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFQAOHLRACQOWLHKQAMQSGSGP 38

RESULT 6
ID Q41196 PRELIMINARY; PRT; 145 AA.
AC Q41196;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE MAJOR ALLERGEN SIN A I (FRAGMENT).
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93151856; PubMed=8093997;
RA Gonzalez de la Pena M.A., Villalba M., Garcia-Lopez J.L.,
RA Rodriguez R.;
RT "Cloning and expression of the major allergen from yellow mustard
RT seeds, Sin a I.";
RL Biochem. Biophys. Res. Commun. 190:648-653(1993).
DR EMBL; S54101; AAB25214.2; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1-tryp_aml_1nh.
DR InterPro; IPR000617; Napin.
DR Pfam; PF00234; try_1alpha_aml1; 1.
DR PRINTS; PR00496; NAPIN.
DR PRODOM; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
FT NON_TER 1 145
SQ SEQUENCE 145 AA; 16449 MW; F8BE8E5CC2C2DDA7 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFQAOHLRACQOWLHKQAMQSGSGP 38

RESULT 7
ID Q41281 PRELIMINARY; PRT; 145 AA.
AC Q41281;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A I.0108 (FRAGMENT).
GN SIN5.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Pena M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodriguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a I.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL; X91798; CAA62908.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1-tryp_aml_1nh.
DR InterPro; IPR000617; Napin.
DR Pfam; PF00234; try_1alpha_aml1; 1.
DR PRINTS; PR00496; NAPIN.
DR PRODOM; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
FT NON_TER 1 145
SQ SEQUENCE 145 AA; 16240 MW; EA0D0E7C46E92C3 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFQAOHLRACQOWLHKQAMQSGSGP 38

RESULT 8
ID Q41277 PRELIMINARY; PRT; 145 AA.
AC Q41277;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A I.0104 (FRAGMENT).
GN SIN1.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Pena M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodriguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a I.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL; X91799; CAA62909.1; -.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Cereal_1-tryp_aml_1nh.
DR InterPro; IPR000617; Napin.
DR Pfam; PF00234; try_1alpha_aml1; 1.
DR PRINTS; PR00496; NAPIN.
DR PRODOM; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
FT NON_TER 1 145
SQ SEQUENCE 145 AA; 16450 MW; 42158FE7D0D78E38 CRC64;

Query Match 38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIQEQONLRQCOEYIKQV--SGQGP 32
DB 10 CRKEFQAOHLRACQOWLHKQAMQSGSGP 38

```

RESULT 9
Q41278      PRELIMINARY;      PRT;      145 AA.
ID 041278
AC 041278:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0105 (FRAGMENT).
GN SIN2.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91800; CA62910.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amy1_inh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16418 MW; 7170F7B8AD78157 CRC64;

Query Match      38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHRLRACQQLHKQAMQSGSP 38

RESULT 10
Q41279      PRELIMINARY;      PRT;      145 AA.
ID 041279
AC 041279:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0106 (FRAGMENT).
GN SIN3.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91801; CA62911.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amy1_inh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16462 MW; 16EB8E5CC2C07F2F CRC64;

Query Match      38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHRLRACQQLHKQAMQSGSP 38

```

```

DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16268 MW; 03158FE7D6CBD521 CRC64;

Query Match      38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHRLRACQQLHKQAMQSGSP 38

RESULT 11
Q41280      PRELIMINARY;      PRT;      145 AA.
ID 041280
AC 041280:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALLERGEN SIN A 1.0107 (FRAGMENT).
GN SIN4.
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RA Gonzalez de la Peña M.A., Monsalve R.I., Batanero E., Villalba M.,
RA Rodríguez R.;
RT "Expression in Escherichia coli of the major allergen from mustard,
RT Sin a 1.";
RL J. Biol. Chem. 237:827-832(1996).
DR EMBL: X91802; CA62912.1; -.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_amy1_inh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_amy1; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 16462 MW; 16EB8E5CC2C07F2F CRC64;

Query Match      38.8%; Score 71; DB 10; Length 145;
Best Local Similarity 44.8%; Pred. No. 0.017;
Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

QY 6 CRGQIQEQNLRCQCEYIKQV--SGQSP 32
DB 10 CRKEFQAQHRLRACQQLHKQAMQSGSP 38

RESULT 12
Q41168      PRELIMINARY;      PRT;      152 AA.
ID 041168
AC 041168:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STORAGE PROTEIN (FRAGMENT).
GN NAPIN.
OS Raphanus sativus (Radish).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Raphanus.
OX NCBI_TaxID=3726;

```


RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-IMMATURE SEED;
 RX MEDLINE-91216448; PubMed-2022325;
 RA Raynal M., Deligny D., Grellier F., Delseny M.;
 RT "Characterization and evolution of napin-encoding genes in radish and
 related crucifers.";
 RL Gene 99:77-86(1991).
 DR EMBL: M63842; AAA63471.1;
 DE InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR SMART: SM00499; AAT; 1.
 FT NON_TER 1 1
 FT CHAIN 7 41
 FT CHAIN 61 150
 SQ SEQUENCE 152 AA; 17459 MW; 9885CACEA6C0755 CRC64;
 STORAGE PROTEIN SMALL SUBUNIT.

Query Match 38.8%; Score 71; DB 10; Length 152;
 Best Local Similarity 44.8%; Pred. No. 0.018;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIOEQNLRCQCEYIKQV--SGQGP 32
 DB 16 CRREFQAOHLRACQOMLHRQARSGSGP 44

RESULT 13
 ID 041169 PRELIMINARY; PRT; 174 AA.
 AC 041169;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE STORAGE PROTEIN (FRAGMENT).
 GN NAPIN.
 OS Raphanus sativus (Radish).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Raphanus.
 OX NCBI_TaxID=3726;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-IMMATURE SEED;
 RX MEDLINE-91216448; PubMed-2022325;
 RA Raynal M., Deligny D., Grellier F., Delseny M.;
 RT "Characterization and evolution of napin-encoding genes in radish and
 related crucifers.";
 RL Gene 99:77-86(1991).
 DR EMBL: M63843; AAA63472.1;
 DE InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR PRODOM: PD002498; Napin.1.
 DR SMART: SM00499; AAT; 1.
 FT NON_TER 1 1
 FT CHAIN 7 41
 FT CHAIN 61 150
 SQ SEQUENCE 174 AA; 19928 MW; 966F0CF7B19A/FPS CRC64;
 STORAGE PROTEIN SMALL SUBUNIT.

Query Match 38.8%; Score 71; DB 10; Length 174;
 Best Local Similarity 44.8%; Pred. No. 0.021;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIOEQNLRCQCEYIKQV--SGQGP 32
 DB 38 CRREFQAOHLRACQOMLHRQARSGSGP 66

RESULT 14
 ID 042469 PRELIMINARY; PRT; 180 AA.
 AC 042469;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE NAPIN.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. TOPAS;
 RX MEDLINE-95161697; PubMed-7858212;
 RA Boullier K.A., Gines M.J., Demoor J.M., Huang B., Baszczynski C.L.,
 Iyer V.N., Miki B.L.;
 RT "Expression of the BmNAP subfamily of napin genes coincides with the
 induction of Brassica microspore embryogenesis.";
 RL Plant Mol. Biol. 26:1711-1723(1994).
 DR EMBL: U04944; AAA81908.1;
 DE InterPro: IPR003612; AAT.
 DR InterPro: IPR001768; Cereals-trypan-amylin.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; trypan-amylin.1.
 DR PRINTS: PR00496; NAPIN.
 DR PRODOM: PD002498; Napin.1.
 DR SMART: SM00499; AAT; 1.
 SQ SEQUENCE 180 AA; 20342 MW; 327FC2A7D9838E3C CRC64;

Query Match 38.8%; Score 71; DB 10; Length 180;
 Best Local Similarity 44.8%; Pred. No. 0.022;
 Matches 13; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

OY 6 CRGQIOEQNLRCQCEYIKQV--SGQGP 32
 DB 47 CRREFQAOHLRACQOMLHRQARSGSGP 75

RESULT 15
 ID 09S9F1 PRELIMINARY; PRT; 41 AA.
 AC 09S9F1;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE NAPIN SHORT CHAIN S4-CALMODULIN ANTAGONIST/CALCIUM-DEPENDENT PROTEIN
 GN KINASE SUBSTRATE.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE-96283790; PubMed-8679670;
 RA Neumann G.M., Condron R., Thomas I., Polya G.M.;
 RT "Purification and sequencing of multiple forms of Brassica napus seed
 napin small chains that are calmodulin antagonists and substrates for
 plant calcium-dependent protein kinase.";
 RL Biochim. Biophys. Acta 1295:23-33(1996).
 SQ SEQUENCE 41 AA; 4652 MW; 8BC6738503380553 CRC64;

Query Match 38.3%; Score 70; DB 10; Length 41;
 Best Local Similarity 41.9%; Pred. No. 0.0064;
 Matches 13; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:42:39 ; Search time 94.6 Seconds
(without alignments)
/6.319 Million cell updates/sec

Title: US-09-913-351-5

Perfect score: 350
Sequence: 1 QERSLRGCCDHLKQMSQCR.....FRTANLPSCGVSPTECRF 65

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:

747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A-Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350	100.0	65	21	AA15423
2	136	38.9	158	18	AAW23586
3	134	38.3	158	18	AAW23588
4	133.5	38.1	111	19	AAW6269
5	132.5	37.9	93	12	AA14111
6	132.5	37.9	110	15	AAW62762
7	132.5	37.9	111	19	AAW6271
8	132.5	37.9	111	19	AAW6264
9	132.5	37.9	140	10	AAW1891
10	131.5	37.6	171	15	AAW3579
11	130.5	37.3	111	19	AAW6267

12	130.5	37.3	111	19	AAW6268	Modified Brazil nu
13	130.5	37.3	111	19	AAW6270	Modified Brazil nu
14	130.5	37.3	149	19	AAW6272	Modified Brazil nu
15	130.5	37.3	149	19	AAW6273	Modified Brazil nu
16	129.5	37.0	111	19	AAW6266	Modified Brazil nu
17	129.5	37.0	111	19	AAW6265	Modified Brazil nu
18	121	34.6	68	21	AA15422	Castor bean 2S alb
19	117.5	33.6	149	19	AAW6274	Modified Brazil nu
20	112	32.0	164	10	AAW1892	Amino acid sequenc
21	112	32.0	164	10	AAW6144	Sequence encoded b
22	108	30.9	148	21	AAW58049	Puroindoline B pro
23	108	30.9	148	21	AAW58050	Serine substituted
24	106.5	30.4	155	18	AAW10569	Mablin II prepro
25	106.5	30.4	155	18	AAW23587	Mablin MBLIT firo
26	104	29.7	166	21	AAW33600	Modified Ara h 2 a
27	104	29.7	166	22	AAW5035	Modified anaphylac
28	104	29.7	167	22	AAW4710	Modified anaphylac
29	101	28.9	156	20	AAW40973	Ara h 2 protein fr
30	101	28.9	157	18	AAW24164	Peanut allergen Ar
31	101	28.9	157	18	AAW24153	Peanut allergen Ar
32	101	28.9	157	20	AAW15245	Peanut allergen A
33	101	28.9	157	22	AAW04707	Anaphylactic anti
34	101	28.9	166	20	AAW40968	Recombinant Ara h
35	101	28.9	207	22	AAW82383	Peanut allergen Ar
36	99	28.3	175	14	AAW33390	Albumin-type prote
37	96.5	27.6	184	12	AAW41493	Napin 1-2 gene pro
38	96.5	27.6	184	15	AAW54981	Napin protein from
39	96.5	27.6	184	15	AAW6694	B. campestris napl
40	96.5	27.6	184	18	AAW15415	Napin encoded by c
41	96.5	27.6	184	19	AAW30526	Brassica campestri
42	96.5	27.6	184	22	AAW85931	B. campestris lamb
43	96.5	27.6	189	9	AAW81140	Sequence encoded b
44	95.5	27.3	184	19	AAW30670	Napin seed storage
45	94.5	27.0	189	15	AAW54980	Napin protein from

ALIGNMENTS

RESULT 1	AA15423	standard; peptide: 65 AA.
ID	AA15423	
AC	AA15423	
DT	17-JAN-2001	(first entry)
DE	Castor bean 2S albumin storage protein amino acids 194-258.	
KW	Antitumor; antimicrobial; immunostimulatory; glycoconjugate; mannose;	
KW	polysaccharide; mannose; galactose; castor bean; immune response; human;	
KW	2S albumin storage protein; animal; tumor necrosis factor; mononuclear;	
KW	phagocyte; granulocyte; macrophage; tumor; immunogen; Candida utilis.	
XX		
OS	Ricinus communis.	
XX		
FN	WO200050087-A1.	
XX		
PD	31-AUG-2000.	
XX		
PF	21-OCT-1999; 99WO-ES00338.	
XX		
PR	26-FEB-1999; 99ES-0000408.	
XX		
PA	(INFA-) IND FARM CANTABRIA SA.	
XX		
PI	Brieva Delgado A, Garcia Villarrubia V, Guerrero Gomez-Pamo A;	
PI	Piñel Ranieri JP, Gimenez Gallego G, Matji Tuduri JA;	
XX		
DR	WPI: 2000-558369/51.	
XX		
PT	New glycoconjugate, useful for treating immunological disorders,	
PT	comprises polysaccharide from Candida utilis and polypeptide from	

PT	Ricinus communis	-	
XX	PS	Claim 3; Page 30; 38pp; Spanish.	
CC	XX	The invention relates to a novel glycoconjugate (A) comprising a	
CC	CC	50-250 kDa polysaccharide (PS), containing one phosphate group per	
CC	CC	5-25 monosaccharide units, and a polypeptide (PP) having a particular	
CC	CC	consensus sequence. In PS, at least 40% of the monosaccharide residues	
CC	CC	are mannose and the remainder are glucose and/or galactose. The main	
CC	CC	chain comprises 1-6 links, with less than 60% 1,2-branching. The PP has	
CC	CC	the consensus sequence:	
CC	CC	Zac2bC(QERK)2(Z')2(LIVM)2cCC(Z'')2(DEH)(LV)26C2C22242g2cZ(VIIM)2FC2g	
CC	CC	where Z = any amino acid (aa); Z' = hydrophobic aa; Z'' = hydrophilic aa;	
CC	CC	a = 3-8; b = 9-13; c = 15-39; d = 13-56; e = 15-26; f = 1-8; g = 1-11;	
CC	CC	the parentheses indicate the preferred sequence. This sequence	
CC	CC	corresponds to amino acids 194-258 of the Ricinus communis (castor bean)	
CC	CC	2S albumin storage protein which is used to generate the glycoconjugate	
CC	CC	of the invention, with a polysaccharide from Candida utilis. (A) modifies	
CC	CC	the immune response in humans and animals, especially it reduces the	
CC	CC	amount of tumour necrosis factor (TNF) produced and stimulates the	
CC	CC	mononuclear-phagocyte system and expands the granulocyte-macrophage	
CC	CC	compartment. (A) are used for treating disorders of the immune system,	
CC	CC	e.g. infections and tumours. (A) do not interfere with metabolic	
CC	CC	processes in the liver and aurs are active when given orally. PP significantly	
CC	CC	increases the antigenic response to weakly immunogenic PS, including	
CC	CC	induction of a T-dependent response.	
XX	Sequence	65 AA:	
SO			
	Query Match	100.0%; Score 350; DB 21; Length 65;	
	Best Local Similarity	100.0%; Pred. No. 2.9e-37;	
	Matches	65; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1	QERSLNGCCDHLKOMOSQCRCEGLRQAIIEQOQSOGLOFGDVEAFRTAANLPSMCVSP	60
DB	1	qerslrgccdhklkmqsgqrceglrqaieqgsgqqlggqdvreaftaanlpmscvsp	60
OY	61	TECRP	65
DB	61	tecrf	65
RESULT	2		
AAW23586			
ID	AAW23586	standard; Protein; 158 AA.	
XX	AAW23586;		
AC			
XX			
DT	30-SEP-1997	(first entry)	
XX			
DE			
XX			
MM	MBL: Sweetener; transgenic plant; recombinant protein; food;		
KW	beverage; animal feed; chewing gum; dental hygiene product;		
KW	pharmaceutical.		
XX			
OS	Capparis masakal.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..39	
FT		/label= Signal	
FT	Protein	40..158	
FT		/label= MBLI	
FT	Region	40..70	
FT		/label= A_chain	
FT	Region	71..86	
FT		/note= "14 amino acids evidently lost when mature	
FT	Region	87..158	
FT		dimer is formed"	
FT		/label= B_chain	
PN	W09700945-A1		

[illegible]

FT	Region	87...158	
FN		/label= B_chain	
XX	W09700945-A1.		
XX	09-JAN-1997.		
PD	21-JUN-1996;	96WO-US10669.	
XX	23-JUN-1995;	95US-0000480.	
XX	(UYHA-) UNIV HAWAII.		
XX	Chen H, Hu Z, Sun SSM, Xiong L;		
XX	WPI: 1997-087372/08.		
DR	Mablinin nucleotide sequences - for prodn. of transgenic plants or		
XX	recombinant proteins, useful as sweeteners, partic. in foods		
PT	Disclosure; Fig 3; 38pp; English.		
XX			
XX	The present sequence represents the mablinin MBLII. The encoding CDNA		
CC	can modulate the expression of the gene encoding MBL in a cell or		
CC	tissue. Mablinin comes from the seeds of the Capparis masaiakal plant,		
CC	and is limited by its native availability, and so methods for producing		
CC	recombinant mablinin protein are being produced using chimeric genes		
CC	such as MBLII (see AAt60776). The products provide recombinant materials		
CC	for the production of the MBL protein in practical amounts, and for the		
CC	production of transgenic plants containing inherently sweet edible parts		
CC	by virtue of production of MBL in situ. The MBL protein can be used to		
CC	sweeten foods, beverages, animal feeds, chewing gum, dental hygiene		
CC	products and pharmaceuticals.		
XX			
SO	Sequence 158 AA;		
Query Match	38.3%; Score 134; DB 18; Length 158;		
Best Local Similarity	43.5%; Pred. No. 3.3e-09;		
Matches 30; Conservative 12; Mismatches 23; Indels 4; Gaps 3;			
OY	1 OER--STGSCDDHLKOMOSQCRCEGLRQAIHQOOSQGLQG-QDYVEAFRTANIPSMCG 57		
Db	87 qqrqpalllccnqrlrvnkpvcvpylrgaahqilyggqlegprvrrlffraainlpnck 146		
OY	58 VSPR-ECRF 65		
Db	147 ipavgrcqv 155		
RESULT 4			
AAW86269			
ID	AAW86269 standard; Protein; 111 AA.		
XX	AAW86269;		
XX	01-MAR-1999 (first entry)		
DE	Modified Brazil nut 2S albumin gene BN18 amino acid sequence.		
XX			
KW	Brazil Nut 2S albumin; seed storage protein; modified; BMCNSS; BN11;		
KW	BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN19; AT2S1BN153W;		
KW	chimeric; dietary protein; essential amino acid.		
XX	Bertholletia excelsa.		
OS	Synthetic.		
XX	W09845458-A1.		
XX	15-OCT-1998.		
XX	06-APR-1998; 98WO-US06673.		
XX			

XX	08-APR-1997.	97US-0042827.
XX		
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	
XX		
PI	Gutteridge S;	
XX		
MP1	1998-609902/51.	
DR	N-PSDB; AAV80249.	
XX		
PT	New modified Brazil Nut 2S albumin proteins - useful to provide	
PT	transgenic seed crops whose proteins have increased levels of	
PT	essential amino acids	
XX		
PS		
XX	ClaIm 6; Fig 6; 80pp; English.	
XX		
CC	The invention relates to modified Brazil Nut 2S albumin seed storage	
CC	proteins and nucleic acids encoding the modified proteins. The modified	
CC	Brazil nut 2S albumin proteins are rich in essential amino acids and the	
CC	encoding genes selected from a group comprising of BNCNS, BN11, BN15,	
CC	BN17, BN18, BN19, BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes.	
CC	Chimeric genes comprising a nucleic acid fragment encoding a modified	
CC	Brazil nut 2S protein operably linked to suitable regulatory sequences,	
CC	integrated into the genome of a transformed plant can be used in	
CC	methods for increasing the essential amino acid content of seeds. The	
CC	products can be used to provide a human and animal dietary protein in	
CC	seed crop plants that is relatively high in essential amino acid content.	
CC	The present sequence represents the amino acid sequence of a modified	
CC	Brazil nut 2S albumin gene BN18. The modified BN18 has all the 15 Arg	
CC	residues, Gly at position 105, and Ser at positions 44 and 107 of the	
CC	wild-type protein replaced with Lys residues	
XX		
XQ		
Sequence	111 AA;	

```

Query Match          38.1%; Score 133.5; DB 19; Length 111;
Best Local Similarity 35.4%; Pred. No. 2.5e-09;
Matches    23; Conservative   17; Mismatches   24; Indels    1; Gaps    1.

QY      2 ERSUNGCCDHLKOMOSOCRCGLR-QAIEQQQSOGQLGQDVFEAFRTAAALPMSGVSP 60
        |::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       40 ephmkececeglegmdesckcglkmmnmknkgqemgpkgyeqmklamnlpskcslsp 99

QY      61 TCRF 65
        :||
Db       100 mkcpf 104

RESULT      5
ID AARI4111 standard; Protein; 93 AA.
XX AARI4111;
XX AC
XX
DT 10-DEC-1991 (first entry)
DE Brazil nut 12 kD sulfur-rich seed storage protein, B2-3.
XX
XX Expression cassette; tryptophan; foodstuff; bovine somatotropin;
KM growth hormone.
XX
XX Synthetic.
XX OS
XX
FH Key Location/Qualifiers
FT Region 24..89
FT Region /label= B2-2
FT Region 28..75
FT Region /label= B2-1
PN WO9113993-A.
XX
XX
PD 19-SEP-1991.
PF 14-FEB-1991; 91WO-US00887.
XX
```

XX	05-MAR-1990;	90US-0489175.
XX	(UPJO)	UPJOHN CO.
XX		
XX	Slightcom JL,	Chee PP;
XX		
XX	WPI: 1991-295647/40.	
XX	N-PSDB; AAQ13801.	
XX		
PT	Seed specific expression cassettes - having specific gene	
PT	regulatory elements for prodn. of proteins in plant seed storage	
PT	bodies	
XX		
PS	Disclosure; Page 38; 58pp; English.	
XX		
CC	Three oligomers were synthesised representing portions of the Brazil	
CC	nut coding DNA. The first oligomer (tag b) encodes the major portion	
CC	of the 9 kd subunit (R2-1), which contains 35% sulfur amino acids for	
CC	the 48 amino acid peptide. The second oligomer (tag a) consisting of	
CC	the complete 9 kd subunit (R2-2) encodes 70 amino acids contg. 28.5%	
CC	sulfur amino acids. The third oligomer is the complete 12 kd subunit	
CC	(R2-3) which encodes 93 amino acids contg. 28% sulfur amino acids.	
CC	The fragments were complemented, purified, and had NcoI sites added on	
CC	both 5' and 3' ends using CPCR amplification. Each inframe NcoI site	
CC	yields the codon TGC which encodes tryptophan, making resulting plant	
CC	seed a more nutritionally complete foodstuff. The fragments were	
CC	subjected to digestion with NcoI and cloned into NcoI cut expression	
CC	cassettes. Plasmids contg. such expression cassettes were transferred	
CC	into vectors used for Agrobacterium systems or used with microprojectile	
CC	bombardment of plant tissues.	
CC	The seed specific expression cassette is used for the prodn. of	
CC	desired proteins in plants, e.g. bovine somatotropin in seed storage	
CC	bodies.	
XX	See also AAQ13798.	
XX		
XX	Sequence 93 AA:	
XX		

Query Match	37.9%	Score 132.5	DB 12	Length 93
Best Local Similarity	39.7%	Pred. No. 2.7e-09		
Matches 25	Conservative 10	Mismatches 27	Indels 1	Gaps 1

QY	2	ERSLGGCDHLKOMOSQCRCEGLR-QAIEQOOSQSGOLQSDVFEAFRTANLPMSGVSP	60
		: : : : : : : :	
Db	29	ephmseeceqlegmdeserccglmmmmmqgemqpgsgmrrmmrlaenlpsrcnlsp	88

QY	61	TEC	63
Db	89	mrc	91

RESULT	6
AA62762	
ID	AA62762 standard; Protein; 110 AA.
XX	
AA62762;	
XX	
AC	
XX	
DT	07-JUL-1995 (first entry)
XX	
DE	Synthetic 12 Kd precursor protein of Brazil nut.
XX	
KW	Stabilisation; toxic; peroxisome; yeast; food additive; brazil nut.
XX	
OS	Synthetic.
XX	
PN	W09424289-A.
XX	
PD	27-OCT-1994:
XX	
PF	19-APR-1994; 94WO-FR00438.
XX	
PR	19-APR-1993; 93FR-0004583.

xx	(EURO-) EUROLXSTINE.
pa	
xx	
pi	Ito H, Iabat N, Micaud J, Pardo D, Raynal A, Sugimoto S;
xx	
dr	WPI: 1994-341868/42.
DR	N-PSDB; AAQ73069.
xx	
pr	Improving stability of protein or peptide in yeast - by
pr	expressing with a peroxisomal targetting sequence, allowing
pr	accumulation of unstable or toxic products, and new yeast strains
pr	useful as food additives
xx	
ps	
xx	Claim 8; Figure 1; 40pp; French.
xx	
cc	Stabilisation of a non-peroxisomal protein or peptide that is
cc	unstable in, or toxic for, a yeast and made by expressing a
cc	heterologous DNA sequence in the yeast, comprises expressing it with
cc	a sequence that targets to peroxisomes to stabilise/detoxify in the
cc	yeast peroxisomes. The method is especially used when the non-
cc	peroxisomal protein is the 2S storage protein (this sequence) of
cc	Brazil nuts or the synthetic peptides E (See AAK62763) and EwE (See
cc	AAK62764). Yeast expressing these proteins are useful as food
cc	additives (having a high content of essential amino acids).
xx	
xx	
SO	Sequence 110 AA;

```

Query Match          37.9%; Score 132.5; DB 15; Length 110;
Best Local Similarity 39.7%; Pred. No. 3.3e-09;
Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1;

OY 2 ERSLEGGCDHLKQWMSQCRCEGLR-QATEEQOSGOLGQDVFEAFPRPANTLPSMCGVSP 60
   | : ||: | : | ||||| : ||: | : | : | | | : ||
Db 41 ephmsceceqlegmdscrcceglrmmmmrmqgeemqpr9eqmrrmmrllaeniprcnisp 100

OY 61 TEC 63
   |
Db 101 mrc 103

```

```

RESULT      7
ID          AAM86271 standard; Protein; 111 AA.
XX
XX          AAM86271;
XX
XX          01-MAR-1999 (first entry)
XX
XX          Modified Brazil nut 2S albumin gene BN153KW amino acid sequence.
DE
XX
XX          Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11,
KW          BN15; BN17; BN18; BN19; BN153KW; AT251BN15; AT251BN15W;
KW          chimeric; dietary protein; essential amino acid.
XX
XX          Bertholletia excelsa.
OS          Synthetic.
OS
XX          Key
FH          Location/Qualifiers
FT          Misc-difference 4
FT          /label= EAK
FT          /note= "wild-type Glu is replaced with Lys"
FT          Misc-difference 6
FT          /label= R6K
FT          /note= "wild-type Arg is replaced with Lys"
FT          Misc-difference 11
FT          /label= R11K
FT          /note= "wild-type Arg is replaced with Lys"
FT          Misc-difference 16
FT          /label= S16K
FT          /note= "wild-type Ser is replaced with Lys"
FT          Misc-difference 19
FT          /label= R19K

```

FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	23
FT		/label= R23K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	36
FT		/label= R36K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	37
FT		/label= R37K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	41
FT		/label= P41K
FT		/note= "wild-type Pro is replaced with Lys"
FT	Misc-difference	42
FT		/label= W42H
FT		/note= "wild-type Trp is replaced with His"
FT	Misc-difference	58
FT		/label= R58K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	63
FT		/label= R63K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	68
FT		/label= R68K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	77
FT		/label= R77K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	82
FT		/label= R82K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	83
FT		/label= R83K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	86
FT		/label= R86K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	94
FT		/label= R94K
FT		/note= "wild-type Arg is replaced with Lys"
FT	Misc-difference	101
FT		/label= R101K
FT		/note= "wild-type Arg is replaced with Lys"
PN	WO9845458-A1.	
PD	15-OCT-1998.	
PP	06-APR-1998;	98WO-US06673.
PR	08-APR-1997;	97US-0042827.
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.	
PI	Gutteridge S;	
DR	WPI: 1998-609902/51.	
DR	N-PSDB; AAV80251.	
PT	New modified Brazil Nut 2S albumin proteins - useful to provide	
PT	transgenic seed crops whose proteins have increased levels of	
PT	essential amino acids	
PS	Claim 6; Fig 8; 80pp; English.	
XX	The invention relates to modified Brazil Nut 2S albumin seed storage	
XX	proteins and nucleic acids encoding the modified proteins. The modified	
XX	Brazil nut 2S albumin proteins are rich in essential amino acids and the	
XX	encoding genes selected from a group comprising of BMCNS, BN11, BN15,	
XX	BN17, BN18, BN19, BN153W, AT251BN15, AT251BN19, AT251BN153W genes.	
XX	Chimeric genes comprising a nucleic acid fragment encoding a modified	
XX	Brazil nut 2S protein operably linked to suitable regulatory sequences,	
XX	integrated into the genome of a transformed plant can be used in	

CC	methods for increasing the essential amino acid content of seeds. The
CC	products can be used to provide a human and animal dietary protein in
CC	seed crop plants that is relatively high in essential amino acid content.
CC	The present sequence represents the amino acid sequence of a modified
CC	Brazil Nut 2S albumin gene BN153KW.
XX	
SQ	Sequence 111 AA:
	Query Match 37.9%; Score 132.5; DB 19; Length 111;
	Best Local Similarity 35.4%; Pred. No. 3.4e-09;
	Matches 23; Conservative 17; Mismatches 24; Indels 1; Gaps 1;
OY	2 ERSLRGCCDHLKQMOSCRCEGLR-QAIEPDQSOGLOGDVFEAFRTAANLPSCGVSF 60
DB	1 : 11 : 1 : 1 : 1111 : :: 11 : 1 : 1 : 1 : 11 : 11 : 11
	40 ekmsmececeglegmdesckceglkmmnmknqgemgpkgyeqnkmklaelnpskcslsp 99
OY	61 TDCRF 65
	: 11
Db	100 mkcpl 104
RESULT	8
AAM86264	
ID	AAM86264 standard; Protein; 111 AA.
XX	
AC	AAM86264;
XX	
DT	01-MAR-1999 (first entry)
XX	
DE	wild-type Brazil nut 2S albumin seed storage protein.
XX	
KM	Brazil Nut 2S albumin; seed storage protein; modified; BNCNSS; BN11;
KW	BN15; BN17; BN18; BN19; BN153KW; AT2S1BN15; AT2S1BN153W;
KX	chimeric; dietary protein; essential amino acid.
XX	
OS	Bertholletia excelsa.
XX	
PN	M09845458-A1.
XX	
PD	15-OCT-1998.
XX	
PF	06-APR-1998; 98WO-US06673.
XX	
PR	08-APR-1997; 97US-0042827.
XX	
PA	(DDPO) DU PONT DE NEMOURS & CO E I.
XX	
PI	Gutteridge S;
XX	
DR	WPI: 1998-609902/51.
N-P	N-PSTDB; AAVB0240.
XX	
PV	New modified Brazil Nut 2S albumin proteins - useful to provide
PT	transgenic seed crops whose proteins have increased levels of
PT	essential amino acids
XX	
PS	Example 1; Fig 1; 80pp; English.
XX	
CC	This represents a wild-type Brazil Nut 2S albumin seed storage
CC	protein. The invention relates to modified Brazil Nut 2S albumin
CC	proteins that are rich in essential amino acids and nucleic acids
CC	encoding the modified proteins. The modified Brazil nut 2S albumin are
CC	selected from a group comprising of BNCNSS, BN11, BN15, BN17, BN19,
CC	BN153KW, AT2S1BN15, AT2S1BN19, AT2S1BN153W genes. Chimeric genes
CC	comprising a nucleic acid fragment encoding a modified Brazil nut 2S
CC	protein operably linked to suitable regulatory sequences, integrated into
CC	the genome of a transformed plant can be used in methods for
CC	increasing the essential amino acid content of seeds. The products can be
CC	used to provide a human and animal dietary protein in seed crop plants
CC	that is relatively high in essential amino acid content.
XX	
SQ	Sequence 111 AA;

```

Query Match Similarity 37.9%; Score 132.5; DB 19; Length 111;
Best Local Similarity 39.7%; Pred No. 3.4e-09;
Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1.

QY      2 ESSLRCCDHLKOMOSQCRCEGLR-QALIEQOSQSGLOGADYFEARTANLPSMGVSP 60
      1 : 11: 1: 1 111111 : 11 : 1:1: 1:11 1:11
Db      40 ephmsecceqlgmdescrcceglmmmmmqgmcpgrgcmrmmrleaprcnslp 99
QY      61 TEC 63
      1
Db      100 mrc 102

RESULT 9
AAP91891
ID AAP91891 standard; protein; 140 AA.
XX
XX AAP91891;
XX
XX 29-APR-1990 (first entry)
XX
XX
XX Part of the sequence of the Brazil nut 2S-albumin as encoded in the
DE PBN2S1 plasmid.
XX
XX 2S-albumin; brazil nut; PBN2S1; storage protein gene;
XX heterologous polypeptide.
XX
XX Brazil nut.
XX
XX
XX Key Location/Qualifiers
XX FH 1..30
XX FT /note="signal peptide"
XX FT 31..38
XX FT /note="mature small subunit"
XX FT 39..43
XX FT /note="processing site"
XX FT 44..136
XX FT /note="mature large subunit"
XX
XX W08903887-A.
XX PN
XX PD 05-MAY-1989.
XX
XX PF 20-OCT-1988; 88WO-EP00944.
XX PR 20-OCT-1987; 87EP-0402348.
XX PA (PLAN-) PLANT GENETIC SYST.
XX
XX Vandekerckhove JS, Krebbers E, Botterman J, Leemans J;
XX WPI: 1989-150783/20.
XX DR N-PSDB: AAN91699.
XX
XX Recombinant DNA expression in plants
XX PT - using modified storage protein genes for expressing
XX PT heterologous polypeptide(s) in the seeds
XX
XX Fig 4; ; 121pp; English.
XX
XX The entire 2S-albumin storage protein precursor including
XX signal peptide. It is to be inserted into plants under the control of
XX a seed-specific promoter and expressed at high levels only or mostly
XX in the seed forming stage and produced mostly in the seeds.
XX
XX Sequence 140 AA;

```

[illegible]

Query Match	37.98;	Score 132.5;	DB 10;	Length 140;
Best Local Similarity	39.78;	Pred. No. 4.4e-09;		
Matches 25;	Conservative 10;	Mismatches 27;	Indels 1;	Gaps 1

Query Match	37.68;	Score 131.5;	DB 15;	Length 171;
Best Local Similarity	44.18;	Pred. No. 7.6e-09;		
Matches 30;	Conservative 14;	Mismatches 17;	Indels 7;	Gaps 4;

• • • • •

```

; Patent No. 5789380
; GENERAL INFORMATION:
; APPLICANT: MIYAZAKI, TOSHIYUKI
; APPLICANT: MORIMOTO, TOSHIHISA
; APPLICANT: MURAYAMA, RYUJI
; APPLICANT: TAKASE, SACHIKO
; APPLICANT: GODA, TOSHINO
; TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
; VISCERAL FAT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,177
; FILING DATE: 14-JAN-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 23445/1996
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1327-055-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-782-177-1

```

```

Query Match 20.9%; Score 73; DB 1; Length 124;
Best Local Similarity 27.7%; Pred. No. 0.047;
Matches 18; Conservative 10; Mismatches 27; Indels 10; Gaps 2;

```

```

QY 2 ERLRGCCDHLKQMOSQCRCEGLRAIEQOOSGGLQ---GQDVEAFR-----TAAN 51
| | | | | : | | | | | : | | | | | : | | | | |
DB 35 EAVLRDCCQQLADISWRCGALYSMLDSMYKHEGVQEGAGTGAFPPCRREVVKLTAA 94
| | | | | : | | | | | : | | | | | : | | | | |
QY 52 LPSCMC 56
: : : : :
DB 95 ITAVC 99

```

```

RESULT 15
US-08-782-177-3
; Sequence 3, Application US/08782177
; Patent No. 5789380
; GENERAL INFORMATION:
; APPLICANT: MIYAZAKI, TOSHIYUKI
; APPLICANT: MORIMOTO, TOSHIHISA
; APPLICANT: MURAYAMA, RYUJI
; APPLICANT: TAKASE, SACHIKO
; APPLICANT: GODA, TOSHINO
; TITLE OF INVENTION: AGENTS FOR INHIBITING ACCUMULATION OF
; VISCERAL FAT
; NUMBER OF SEQUENCES: 5

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,177
; FILING DATE: 14-JAN-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 23445/1996
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1327-055-0X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-782-177-3

```

```

Query Match 20.9%; Score 73; DB 1; Length 124;
Best Local Similarity 29.2%; Pred. No. 0.047;
Matches 19; Conservative 9; Mismatches 27; Indels 10; Gaps 2;

```

```

QY 2 ERLRGCCDHLKQMOSQCRCEGLRAIEQO-OSGGLQGGDVEAF-----TAAN 51
| | | | | : | | | | | : | | | | | : | | | | |
DB 35 EAVLRDCCQQLADISWRCGALYSMLDSMYKHEGVQEGAGTGAFPPCRREVVKLTAA 94
| | | | | : | | | | | : | | | | | : | | | | |
QY 52 LPSCMC 56
: : : : :
DB 95 ITAVC 99

```

```

Search completed: September 9, 2002, 12:44:14
Job time: 236 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:43:30 ; Search time 45.25 Seconds
(without alignments)
138.029 Million cell updates/sec

Title: US-09-913-351-5

Sequence: 1 QBRSLRGCCDHLKQMOSQCR.....FRITANLPSCGVSPTGCRF 65

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_71:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350	100.0	258	1	RZCS
2	155.5	44.4	70	2	seed storage pr
3	147	42.0	154	2	seed storage prote
4	135	38.6	104	2	albumin - Brazil
5	133	38.0	100	2	albumin I-1 - Yun
6	133	38.0	104	2	albumin IV - Yun
7	132.5	37.9	146	2	albumin III - Yun
8	130	37.1	141	2	albumin precurs
9	119.5	34.1	295	2	seed storage pr
10	113.5	32.4	165	2	seed storage prote
11	112.5	32.1	139	2	albumin 2S storage
12	112.5	32.1	139	2	albumin 2S storage
13	112	32.0	164	1	albumin 2S storage
14	110.5	31.6	323	2	albumin I precu
15	108	30.9	148	2	albumin - commo
16	106.5	30.4	72	2	albumin II chain
17	106.5	30.4	155	2	albumin II precu
18	104.5	29.9	178	1	albumin 2 precursor
19	104.5	29.9	178	2	albumin B - rape
20	104.5	29.9	178	2	albumin precursor (n
21	104.5	29.9	178	2	2S storage protein
22	104	29.7	162	2	2S seed storage pr
23	104	29.7	172	2	2S seed storage pr
24	102.5	29.3	178	2	2S-like storage pr
25	102	29.1	124	2	napin napb - rape
26	101.5	29.0	155	2	napin - Swedish tu
27	101.5	29.0	178	2	napin Bg9 precurs
28	100.5	28.7	158	2	2S storage protein
29	99.5	28.4	162	2	major allergen lar
					2S seed storage pr

30	99	28.3	162	2	S49259	albumin 4 - easter
31	98.5	28.1	186	2	A23602	napin precursor (g
32	98	28.0	152	2	PS0427	napin AH1 precurs
33	97	27.7	174	2	PS0425	napin BA3 precurs
34	96.5	27.6	180	2	S10018	napin (clone BgNA
35	96.5	27.6	180	2	S52025	napin (clones BNM
36	96.5	27.6	327	2	JS0402	gamma-gliadin prec
37	95.5	27.3	155	2	S51770	grain softness pro
38	95.5	27.3	164	2	S48186	grain softness pro
39	95	27.1	145	2	PC1247	Sin a I allergen 2
40	95	27.1	166	1	NMMU4	2S albumin 4 precu
41	94.5	27.0	164	2	S48187	grain softness pro
42	94.5	27.0	170	1	NMMU2	2S albumin 2 precu
43	92	26.3	80	2	B23617	conglutinin delta-2
44	91.5	26.1	186	1	WMR219	19k globulin precu
45	91.5	26.1	186	2	JC4784	alpha-globulin pre

ALIGNMENTS

RESULT 1

RZCS

2S seed storage protein precursor - castor bean

N:Alternate names: 2S albumin precursor

C:Species: Ricinus communis (castor bean)

C:Date: 14-Nov-1983 #sequence_revision 08-Feb-1996 #text_change 18-Jun-1999

C:Accession: S11499; S11500; S11501; S11502; S27221; A01328; S27222

R:Irwin, S.D.; Lord, J.M.

Nucleic Acids Res. 18, 5890, 1990

A:Title: Nucleotide sequence of a Ricinus communis 2S albumin precursor gene.

A:Reference number: S11499; MUID:91016940

A:Accession: S11499

A:Molecule type: DNA

A:Residues: 1-258 <IRW>

A:Cross-references: EMBL:X54158; NID:921067; PIDN:CA38097.1; PID:921068

A:Note: the authors translated the codon CTC for residue 14 as Phe, CCA for residue 7

R:Irwin, S.D.; Keen, J.N.; Findlay, J.B.C.; Lord, J.M.

Mol. Gen. Genet. 222, 400-408, 1990

A:Title: The Ricinus communis 2S albumin precursor: a single preproprotein may be pro

A:Reference number: S11500; MUID:91109729

A:Accession: S11500

A:Molecule type: mRNA

A:Residues: 1-13, 'F', 15-73, 'T', 75-258 <IR2>

A:Experimental source: clone 1494

A:Accession: S11501

A:Molecule type: mRNA

A:Residues: 'M', 4, 'LS', 7-13, 'F', 15-21 <IRF>

A:Experimental source: clone 10a12

A:Accession: S27221

A:Molecule type: protein

A:Residues: 'X', 37, 'X', 39-45; 'X', 158-161, 'X', 163-174, 'X' <IRA>

R:Sharf, F.S.; Li, S.S.L.

J. Biol. Chem. 257, 14753-14759, 1982

A:Title: Amino acid sequence of small and large subunits of seed storage protein from

A:Reference number: A92357; MUID:83082772

A:Accession: A01328

A:Molecule type: protein

A:Residues: 157-190; 194-221, 'Q', 223-225, 230-233, 'N', 235-254, 'Q', 256-258 <SHA>

A:Note: 230-Ser was also found

A:Note: there is considerable similarity between residues 181-231 of this protein and

R:Odani, S.; Koide, T.; Ono, T.; Ohnishi, K.

Biochem. J. 213, 543-545, 1983

A:Title: Structural relationships between barley (Hordeum vulgare) trypsin inhibitor a

A:Reference number: A90322; MUID:83308577

A:Contents: annotation

A:Note: this protein is homologous with trypsin inhibitor from barley

C:Complex: consists of two chains linked by two disulfide bonds involving Cys-162 and

C:Superfamily: 2S seed storage protein precursor

C:Keywords: pyroglutamic acid; seed; storage protein
 F:1-21/Domain: signal sequence #status predicted <STIG>
 F:22-35/Domain: propeptide #status predicted <PRO>
 F:36-72/Product: probable 2S seed storage protein small chain 2 #status experimental <SC>
 F:87-156/Product: probable 2S seed storage protein large 2 #status predicted <LCH>
 F:157-199/Product: 2S seed storage protein small chain #status experimental <SM>
 F:194-258/Product: 2S seed storage protein large chain #status experimental <LRG>
 F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 100.0%; Score 350; DB 1; Length 258;
 Best Local Similarity 100.0%; Pred. No. 1,4e-32;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QRSRLRGCCDHLKQKQMSQCRCEGLROAIEQOOSQOLGQDVFEAFRTANLPSMCGVSP 60
 DB 194 QRSRLRGCCDHLKQKQMSQCRCEGLROAIEQOOSQOLGQDVFEAFRTANLPSMCGVSP 253

OY 61 TECRF 65
 DB 254 TECRF 258

RESULT 2
 A59346 seed storage protein Lec2SA1 large chain [imported] - tomato (fragment)
 N:Alternate names: 2S albumin large subunit
 C:Species: Lycopersicon esculentum (tomato)
 C:Date: 31-Dec-2001 #sequence_revision 31-Dec-2001 #text_change 31-Dec-2001
 C:Accession: A59346
 R:Oguri, S.
 submitted to the Protein Sequence Database, September 2000
 A:Reference number: A59346
 A:Accession: A59346
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-70 <OGU>
 A:Experimental source: cv. cherry; seed
 A:Note: cross-reactive with anti-tomato lectin antiserum: one of four isoforms in seeds
 C:Complex: heterodimer of large and small (see PIR:B59346) chains, disulfide linked
 C:Keywords: heterodimer; pyroglutamic acid; seed; storage protein
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 44.4%; Score 155.5; DB 2; Length 70;
 Best Local Similarity 41.9%; Pred. No. 6.4e-11;
 Matches 26; Conservative 17; Mismatches 18; Indels 1; Gaps 1;

OY 3 RSLRGCCDHLKQKQMS-QCRCEGLROAIEQOOSQOLGQDVFEAFRTANLPSMCGVSP 61
 DB 5 RLVQCCQCNQLEIDPDQCRCEGLRMKVVOEEQGTGRVGROROOMLOTAEINLPGLRLSPQ 64

OY 62 EC 63
 DB 65 RC 66

RESULT 3
 S14947
 2S albumin - Brazil nut
 C:Species: Bertholletia excelsa (Brazil nut)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S14947
 R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross
 Plant Mol. Biol. 16, 437-448, 1991
 A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin fr
 A:Reference number: S14946; MUID:91370890
 A:Accession: S14947
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-154 <GAN>
 A:Cross-references: EMBL:X54491; NID:g17712; PIDN:CAA38363.1; PID:g17713
 C:Genetics:

A:introns: 67/3
 C:Superfamily: wheat alpha-amylase inhibitor

Query Match 42.0%; Score 147; DB 2; Length 154;
 Best Local Similarity 41.3%; Pred. No. 1.3e-09;
 Matches 26; Conservative 12; Mismatches 25; Indels 0; Gaps 0;

OY 1 QRSRLRGCCDHLKQKQMSQCRCEGLROAIEQOOSQOLGQDVFEAFRTANLPSMCGVSP 60
 DB 83 EEPHLDCECCOLEMDEKRCBGLNMLRQREEMELGEBQMRIMRAENLLSCNISP 142

OY 61 TEC 63
 DB 143 QRC 145

RESULT 4
 S48176
 mablinin I-1 - Yunnan caper (fragments)
 C:Species: Capparis masakal (Yunnan caper)
 C:Date: 15-Jul-1995 #sequence_revision 13-Mar-1998 #text_change 13-Mar-1998
 C:Accession: S48176; S48177
 R:Nirasawa, S.; Nishino, T.; Katabira, M.; Uesugi, S.; Hu, Z.; Kurihara, Y.
 Eur. J. Biochem. 223, 989-995, 1994
 A:Title: Structures of heat-stable and unstable homologues of the sweet protein mablin
 A:Reference number: S48176; MUID:94333405
 A:Accession: S48176
 A:Molecule type: protein
 A:Residues: 1-32;33-104 <NIR>
 A:Experimental source: seed
 A:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: albumin; storage protein; sweet taste
 F:1-32/Product: mablinin I-1 chain A (fragment) #status experimental <CHA>
 F:1-32,33-104/Product: mablinin I-1 (fragment) #status experimental <MAT>
 F:33-104/Product: mablinin I-1 chain B #status experimental <HB>
 F:4-53,17-42,43-91,55-99/Disulfide bonds: #status experimental

Query Match 38.6%; Score 135; DB 2; Length 104;
 Best Local Similarity 43.8%; Pred. No. 2.1e-08;
 Matches 28; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

OY 4 SLRGCCDHLKQKQMSQCRCEGLROAIEQOOSQOLGQDVFEAFRTANLPSMCGVSP 61
 DB 38 ALRLCCNQNLQVKNRCVCVPLRQAHQQLYQSIGRPVQLFRANLNLNICKIPAVG 97

OY 62 ECRF 65
 DB 98 RCQF 101

RESULT 5
 S48180
 mablinin IV - Yunnan caper (fragments)
 C:Species: Capparis masakal (Yunnan caper)
 C:Date: 15-Jul-1995 #sequence_revision 13-Mar-1998 #text_change 20-Mar-1998
 C:Accession: S48180; S48181
 R:Nirasawa, S.; Nishino, T.; Katabira, M.; Uesugi, S.; Hu, Z.; Kurihara, Y.
 Eur. J. Biochem. 223, 989-995, 1994
 A:Title: Structures of heat-stable and unstable homologues of the sweet protein mablin
 A:Reference number: S48176; MUID:94333405
 A:Accession: S48180
 A:Molecule type: protein
 A:Residues: 1-28;29-100 <NIR>
 A:Experimental source: seed
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: albumin; storage protein; sweet taste
 F:1-28/Product: mablinin IV chain A (fragment) #status experimental <CHA>
 F:1-28,29-100/Product: mablinin IV (fragment) #status experimental <MAT>
 F:29-100/Product: mablinin IV chain B #status experimental <HB>
 F:4-49,17-38,39-87,51-95/Disulfide bonds: #status predicted

A:Residues: 26-45;65-84 <HAR2>
C:Keywords: seed; storage protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-35/Domain: propeptide #status experimental <PRO>
F:36-141/Product: 2S albumin #status experimental <MAT>

Query Match	37.1%;	Score 130;	DB 2;	Length 141;
Best Local Similarity	39.7%;	Pred. No. 1e-07;		
Matches	23;	Conservative	11;	Mismatches 22;
			Indels	2;
			Gaps	1;

```

Oy      8 CCDHLKOWQSOCCREGLRQAEQOOSQGLQGDVFEAFRTAANLPSMCGVSPTECRF 65
        ||  || :  ||| :  || :  || :  ||| ||| ||| ||| ||| ||| |||
Db      86 CCRELKANDDEECRCDMLEETAREEQRRA--RGQEGRQMLQKARNLPSMCGIRPQRCDF 141

```

RESULT 9
S01062
2S seed storage protein precursor (clone HaG5) - common sunflower

N:Alternate names: 2S albumin storage protein
C:Species: Helianthus annuus (common sunflower)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 21-Jul-2000

A:Reference number: S01062; MUID:88142538
Mol. Gen. Genet. 210, 211-218, 1987
A:Title: Sequence and expression of a gene encoding an albumin storage protein in sunfish

A:Residues: 1-295 <ALT>
A:Cross-references: EMBL:X06410; PTDN:CAA29699.1; PID:g18800
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed correct.

C:introns: 192/2
A:Superfamily: gliadin
F1-20/Domain: signal sequence #status predicted <SIG>
F1-162-195/Product: 2S seed storage protein #status exam

Query Match	34.1%	Score 119.5	DB 2	Length 295
Best Local Similarity	30.3%	Pred. No. 3.5e-06		
Matches 23	Conservative 17	Mismatches 23	Indels 13	Gaps 2

```
QY      1 QERSLRG-----CCDHLKUMOSOCRCCEGLROAI-----EODSOGLOGDVDVEAFR 47
        |::|   |: :: :||| ::|       ||: ||| |||:|
Db    215 QQQQQRGLQQQCNCNELNVKRECHCEALQEVARVMRPDPQQQ0000RRGQFGQGEMETARR 274
```

```
QY      48  TAANLPSCGVSPTEC 63
          |||: |: :|
Db      275  VIQLNPQCDEYQQC 29
```

RESULT 10
T09252
seed storage protein EMB25 - white spruce

C:Accession: T09252
R: Dong, J.Z.; Dunstan, D.I.
submitted to the EMBL data Library, June 1996

A:Accession: T09252
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

C;Genetics:
A;Gene: EMB25

Query Match	32.48;	Score 113.5;	DB 2;	Length 165;
Best Local Similarity	30.78;	Pred. No. 9.4e-06;		

Matches	23;	Conservative	14;	Mismatches	19;	Indels	19;	Gaps	2;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

```

QY      8 CDDHKKOMQSCRCCEGLRQAE-----QQQSGQLQGQDVEEAFRTAA 50
      11 ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      76 CCEELQRMSPCCRCAQTQLTEEDVFMDSQDQAPLNQRRQRRQGRGHEEVEVRAE 135

```

```
QY      51  NLPSCGV--SPTec 63
          ||: || || || |
Db      136  ELPNTCNVRQSPRC 150
```

RESULT	11
T09850	
albumin 2S storage protein precursor - upland cotton	

C:Accession: T09850
R:Galau, G.A.; Wang, H.Y.C.; Hughes, D.W.
submitted to the EMBL Data Library, January 1992

A:Description Cotton Mats (C164) gene and cDNAs encoding a methionine-rich 2S albumin
A:Reference number: Z16886
A:Accession: T09850

A;Residues: 1-139 <GAL>
A;Cross-references: EMBL:M83301; NID:g167310; PID:g167311
C;Genetics:

F;21-139/Product: albumin 2S storage protein #status predicted <MAT>

8 GNNH KQWSSGCGECI DDAITPQQSO--L--CGLACGDVREAFETATN PSMQCIVSPPE 63
Matches 22; Conservative 11; Mismatches 21; Indels 7; Gaps 2;

63 C 63
—
07

RESULT 12

albumin 2S storage protein precursor Mat5-A - upland cotton
C/Species: Gossypium hirsutum (upland cotton)
C/Date: 16-Jul-1999 #sequence revision 16-Jul-1999 #text_ch

submitted to the EMBL Data Library, January 1992
 A:Description: Cotton Mat5-A (Cl64) gene and Mat5-D cDNAs encoding methionine-tRNA 2S
 A:Reference number: Z16893

A;Molecule type: DNA
A;Residues: 1-139 <GAL>
A;Cross-references: EMBL:M86213; NID:q167358; PID:q167359

C:Keywords: storage protein
F1-20/Domain: signal sequence #status predicted <SIG>
F21-139/Product: albumin 2S storage protein Mat5-A #status predicted <MAT>

Query Match	32.1%;	Score 112.5;	DB 2;	Length 139;
Best Local Similarity	39.0%;	Pred. No. 1e-05;		
Matches 23; Conservative	10;	Mismatches 23;	Indels 3;	Gaps 2;

```

0Y      8 CCDHLKQMSQCEGLRAIEQ--QOSQGLQGDVFEAF-RTAANI1PSMGSPT2EC 63
      || ::||:||||||| | ||||: ::| : |||: |||

```


Db 73 CCQGLEKMDTCRCQGLRHATMDQMOMQOMGSKOMREIMQVTKKIMSECEMEPEGRG 131

RESULT 13

NRMT1

2S albumin 1 precursor - Arabidopsis thaliana
N:Alternate names: seed storage protein AT2S1

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 03-Dec-1999

C:Accession: JAO161: PS0282; S34676: T06044

R:Kreberg, E.; Hertles, L.; de Clercq, A.; Seurinck, J.; Leemans, J.; Van Damme, J.; Se

Plant Physiol. 87, 859-866, 1988

A:Title: Determination of the processing sites of an Arabidopsis 2S albumin and character

A:Reference number: JAO161

A:Accession: JAO161

A:Molecule type: DNA

A:Residues: 1-164 <KRE>

A:Cross-references: GB:M22032; NID:g166609; PIDN:AAA32743.1; PID:g166614

A:Accession: PS0282

A:Molecule type: protein

A:Residues: 38-73:84-162 <KR2>

R:Conceicao, A.D.S.; Kreberg, E.

submitted to the EMBL Data Library, July 1993

A:Description: Tentative title: a cotyledon regulatory region is responsible for the dif

A:Reference number: S34674

A:Accession: S34676

A:Molecule type: DNA

A:Residues: 1-164 <CON>

A:Cross-references: EMBL:224745; NID:g395203; PIDN:CAA80870.1; PID:g395204

R:Hevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15484

A:Accession: T06044

A:Molecule type: DNA

A:Residues: 1-164 <BEV>

A:Cross-references: EMBL:AL035680; GSPDB:GN00062; ATSP:T24A18.90

A:Experimental source: cultivar Columbia; BAC clone T24A18

C:Genetics:

A:Gene: T24A18.90

A:Map position: 4

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: seed; storage protein

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-164/Product: 2S albumin 1 proprotein #status predicted <ANT>

F:84-73/Product: 2S albumin 1 small chain #status experimental <SMC>

F:84-162/Product: 2S albumin 1 large chain #status experimental <LAC>

Query Match 32.0%; Score 112; DB 1; Length 164;

Best Local Similarity 35.8%; Pred. No. 1.4e-05;

Matches 24; Conservative 16; Mismatches 25; Indels 2; Gaps 2;

OY 1 QENSLSRGCCDHLKQKQSCRCBGLROAIEQQSQGLOGQDVFPAFRANLPSMGVS 59

Db 89 QEQQLFOCCNEHROBPCVPTLKQAKAVLQGHQHPQYRKIKYQTAKHLPNCVDP 148

OY 60 PTE-CRF 65

Db 149 QVDVCPF 155

RESULT 14

S38887

2S albumin - common sunflower (fragment)

C:Species: Helianthus annuus (common sunflower)

C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-May-2000

C:Accession: S38887

R:Thoyts, P.J.E.; Millichip, M.; Stobart, A.K.; Griffiths, W.T.; Napier, J.A.; Shewry, F

submitted to the EMBL Data Library, November 1993

A:Reference number: S38887

A:Accession: S38887

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-323 <THO>
A:Cross-references: EMBL:X76101; NID:g429181; PID:g429182
C:Superfamily: gliadin

Query Match

Best Local Similarity 31.6%; Score 110.5; DB 2; Length 323;

Matches 21; Conservative 15; Mismatches 27; Indels 5; Gaps 1;

OY 1 QENSLSRGCCDHLKQKQSCRCBGLROAIEQQSQGLOGQDVFPAFRANLPSMGVS 55

Db 252 EQQGLQCCNEHROBPCVPTLKQAKAVLQGHQHPQYRKIKYQTAKHLPNCVDP 311

OY 56 CGVSPTFC 63

Db 312 CDLDVQCC 319

RESULT 15

S46514

putridoline-b precursor - wheat

C:Species: Triticum aestivum (common wheat)

C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Aug-1999

C:Accession: S46514

R:Gautier, M.F.; Aleman, M.E.; Guirao, A.; Marion, D.; Joudrier, P.

Plant Mol. Biol. 25, 43-57, 1994

A:Title: Triticum aestivum putridolines, two basic cysteine-rich seed proteins: cDNA

A:Reference number: S46514; M01D:94272013

A:Accession: S46514

A:Molecule type: mRNA

A:Residues: 1-148 <GAU>

A:Cross-references: EMBL:X69912; NID:g509087; PIDN:CAA49537.1; PID:g509088

C:Superfamily: glycine

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-145/Product: putridoline-a #status predicted <MAT>

F:146-148/Domain: carboxyl-terminal propeptide #status predicted <PRO>

Query Match

Best Local Similarity 30.9%; Score 108; DB 2; Length 148;

Matches 23; Conservative 12; Mismatches 17; Indels 12; Gaps 3;

OY 8 CCHLTKQKQSCRCBGLROAIEQQSQGLOGQDVFPAFRANLPSMGVSPT 61

Db 85 CCKQLSIAPOCKRCDISIRVY-----QGRUGGFLGTWRGEVYQLODPAQSLPSKCMNG-A 138

OY 62 ECRF 65

Db 139 DCKF 142

Search completed: September 9, 2002, 12:43:31
Job time: 223 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:46:06 ; Search time 22.97 Seconds
(without alignments)
109.568 Million cell updates/sec

Title: US-09-913-351-5

350

Perfect score: 1 QERSLRCCCHIKQMOSCR.....FRANLPSCMGVPTPCRF 65

Sequence: BLOSUM62

Gapop 10.0 , Gapext 0.5

Scoring table: 105224 seqs, 38719550 residues

Searched: Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350	100.0	258	1	2SS_RICCO
2	350	100.0	258	1	P01089 ricinus com
3	133	38.6	104	1	P80351 caparis ma
4	133	38.0	100	1	2SS4_CAPMA
5	133	38.0	104	1	2SS3_CAPMA
6	132.5	37.9	146	1	2SS_BEREX
7	112.5	34.1	295	1	2SS5_HELAN
8	112	32.0	164	1	2SS1_ARATH
9	106.5	30.9	148	1	P01090 brassica na
10	106.5	30.4	155	1	2SS2_CAPMA
11	104.5	29.9	178	1	2SS2_BRANA
12	102.5	29.3	178	1	2SSB_BRANA
13	102	29.1	130	1	ITRY_SINAR
14	100.5	28.7	129	1	ALIL_BRATU
15	98.5	28.1	186	1	2SS5_BRANA
16	96.5	27.6	180	1	2SS3_BRANA
17	96.5	27.6	166	1	2SS2_WHEAT
18	95	27.1	327	1	2SS4_ARATH
19	94.5	27.0	170	1	2SS2_ARATH
20	92	26.3	80	1	CG2L_LUPAN
21	91.5	26.1	186	1	GLI9_ORYSA
22	89.5	25.6	60	1	MO2X_MOROL
23	89.5	25.6	127	1	ALIL_SINAR
24	88.5	25.3	164	1	2SS3_ARATH
25	88	25.1	123	1	GLU2_MAIZE
26	86.5	24.7	133	1	2SS1_MAIZE
27	86	24.6	144	1	ITAE_HORYU
28	84.5	24.1	149	1	PRO7_ORYSA
29	83.5	23.9	148	1	PUTA_WHEAT
30	82.5	23.6	305	1	HOG1_HORYU
31	82	23.4	110	1	2SS1_BRANA
32	80	22.9	123	1	ITAE2_WHEAT
33	79	22.6	286	1	GDA0_WHEAT

34	79	22.6	356	1	GLTA_WHEAT	P10385 triticum ae
35	75	21.4	307	1	GDA9_WHEAT	P18573 triticum ae
36	75	21.3	307	1	GLTB_WHEAT	P10386 triticum ae
37	74.5	21.3	158	1	2SS_SOYBN	P19594 glycine max
38	73.5	21.0	117	1	NL75_ORYSA	O65091 oryza sativ
39	73.5	21.0	296	1	GDA6_WHEAT	P04726 triticum ae
40	73.5	21.0	297	1	GDA4_WHEAT	P04724 triticum ae
41	73	20.9	124	1	ITAE1_WHEAT	P01085 triticum ae
42	73	20.9	214	1	AVEN_AVEA	F27919 avena sativ
43	73	20.9	291	1	GDBB_WHEAT	P06659 triticum ae
44	72	20.6	156	1	PRO2_ORYSA	P17048 oryza sativ
45	72	20.6	244	1	GDB3_WHEAT	P04730 triticum ae

ALIGNMENTS

RESULT 1
ID 2SS_RICCO STANDARD: PRT; 258 AA.
AC P01089;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1990 (Rel. 16, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE 2S albumin precursor (Allergen Ric c 1).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Malpighiales; Euphorbiales; Ricinus.
OX NCBI_TaxID=3968;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91109729; PubMed=2274038;
RA Irwin S.D., Keen J.N., Findlay J.B.C., Lord J.M.;
RT "The Ricinus communis 2S albumin precursor: a single preproprotein
Mol. Gen. Genet. 222:400-408(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=91016940; PubMed=2216785;
RA Irwin S.D., Lord J.M.;
RT "Nucleotide sequence of a Ricinus communis 2S albumin precursor
gene.";
RL Nucleic Acids Res. 18:5890-5890(1990).
RN [3]
RP SEQUENCE OF 157-190 AND 194-258.
RX MEDLINE=83082772; PubMed=7174664;
RA Sharief F.S., Li S.S.-L.;
RT "Amino acid sequence of small and large subunits of seed storage
protein from Ricinus communis.";
RL J. Biol. Chem. 257:14753-14759(1982).
RN [4]
RP SIMILARITY TO PROTEINASE INHIBITORS.
RX MEDLINE=83308577; PubMed=6515448;
RA Odani S., Koide T., Ono T., Ohnishi K.;
RT "Structural relationship between barley (Hordeum vulgare) trypsin
inhibitor and castor-bean (Ricinus communis) storage protein.";
RL Biochem. J. 213:543-545(1983).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
LINKED BY 2 DISULFIDE BONDS.
CC -1- PTM: THE TWO DISULFIDE BONDS LINKING THE SMALL AND THE LARGE
CHAINS INVOLVE CYS-162 AND CYS-175.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib.ch).

CC -----
DR EMBL: X54158; CAA38097.1; -
DR PIR: A01328; RZCS.
DR PIR: S11499; S11499.
DR PIR: S11500; S11500.
DR PIR: S11501; S11501.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_1768; Cereals_1768.
DR Pfam: PF00234; trypt_alpha_amyl_2.
DR SMART: SM00499; AAI; 2.
KM Seed storage protein; Signal; Allergen.
FT SIGNAL 1 21
FT PROPEP 22 156
FT CHAIN 157 190 2S ALBUMIN, SMALL CHAIN.
FT PROPEP 191 193
FT CHAIN 194 258 2S ALBUMIN, LARGE CHAIN.
FT MOD_RES 194 194 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 222 222 E -> Q (IN REF. 3).
FT CONFLICT 226 229 D -> N (IN REF. 3).
FT CONFLICT 234 234 E -> Q (IN REF. 3).
FT CONFLICT 255 255
SQ SEQUENCE 258 AA; 29290 MW; 27874CFC50E41072 CRC64;

Query Match 100.0%; Score 350; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.8e-33;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QERSLRGCCDHLKQKQMSQRCRGLRQATEEQOQSGQLOGQDVFEAFRTAANLPSMCGVSP 60
DB 194 QERSLRGCCDHLKQKQMSQRCRGLRQATEEQOQSGQLOGQDVFEAFRTAANLPSMCGVSP 253
OY 61 TECRF 65
DB 254 TECRF 258

RESULT 2
ID 2SS4_CAPMA STANDARD; PRT; 104 AA.
AC P80351.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin I-1, A and B chains (MAB I) (Sweet protein).
OS Capparis masaiikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Capparidaceae;
OX NCBI_TaxID=13395;
RN [1]
RP TISSUE=Seed;
RC MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katsuhira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: 2S SEED STORAGE PROTEIN HAVING SWEETNESS-INDUCING
CC ACTIVITY. THIS FORM IS NOT HEAT STABLE.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_25; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KM Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 32 MABINLIN I, A CHAIN.

FT NON_CONS 32 33
FT CHAIN 33 104 MABINLIN I, B CHAIN.
FT DISULFID 4 53
FT DISULFID 17 42
FT DISULFID 43 91
FT DISULFID 55 99
SQ SEQUENCE 104 AA; 12284 MW; 156A4A4B50D5EE60 CRC64;

Query Match 38.6%; Score 135; DB 1; Length 104;
Best Local Similarity 43.8%; Pred. No. 7.5e-09;
Matches 28; Conservative 12; Mismatches 22; Indels 2; Gaps 2;

OY 4 SLRCCDHLKQKQMSQRCRGLRQATEEQOQSGQLOGQDVFEAFRTAANLPSMCGVSP 61
DB 38 ALRCCNQLRQVKNPCVCVLRQAHAQQLYQGIIEBGRVQLFRARNLNPKICKIPAVG 97
OY 62 ECRF 65
DB 98 RCF 101

RESULT 3
ID 2SS4_CAPMA STANDARD; PRT; 100 AA.
AC P80353;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin IV, A and B chains (MAB IV) (Sweet protein).
OS Capparis masaiikai (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Capparidaceae;
OX NCBI_TaxID=13395;
RN [1]
RP TISSUE=Seed;
RC MEDLINE=94333405; PubMed=8055976;
RA Nirasawa S., Nishino T., Katsuhira M., Uesugi S., Hu Z., Kurihara Y.;
RT "Structures of heat-stable and unstable homologues of the sweet
RT protein mabinlin. The difference in the heat stability is due to
RT replacement of a single amino acid residue.";
RL Eur. J. Biochem. 223:989-995(1994).
CC -1- FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_25; 1.
DR PRINTS: PR00496; NAPIN.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KM Seed storage protein; Albumin; Sweet-taste.
FT CHAIN 1 28 MABINLIN IV, A CHAIN.
FT NON_CONS 28 29
FT CHAIN 29 100 MABINLIN IV, B CHAIN.
FT DISULFID 4 49
FT DISULFID 17 38
FT DISULFID 39 87
FT DISULFID 51 95
SQ SEQUENCE 100 AA; 11928 MW; 7B09673FAB7793CA CRC64;

Query Match 38.0%; Score 133; DB 1; Length 100;
Best Local Similarity 43.8%; Pred. No. 1.2e-08;
Matches 28; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

OY 4 SLRCCDHLKQKQMSQRCRGLRQATEEQOQSGQLOGQDVFEAFRTAANLPSMCGVSP 61
DB 34 ALRCCNQLRQVKNPCVCVLRQAHAQQLYQGIIEBGRVQLFRARNLNPKICKIPAVG 93

Query Match 37.9%; Score 132.5; DB 1; Length 146;
Best Local Similarity 39.7%; Pred. No. 2e-08;
Matches 25; Conservative 10; Mismatches 27; Indels 1; Gaps 1;

QY 2 ERSIRCCDHLKOMQSCRCRGLR-QAIEQSQSGQLQGDVEAFRTAANLPSMCGVSP 60
DB 75 EPHMSSECCEOLEGMDSCRCRGLRMMRMQCEMQPGRGMRMRRLAENIPSCMLSP 134
QY 61 TEC 63
DB 135 MRC 137

RESULT 6
ID 2SS5_HELAN STANDARD; PRT; 295 AA.
AC P15461;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 2S seed storage protein precursor (2S albumin storage protein).
GN HAG5.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 162-173.
RC STRAIN=CV. GIANT GREY STRIPE;
RX MEDLINE=88142538; PubMed=2830455;
RA Allen R.D., Cohen E.A., Vonder Haar R.A., Adams C.A., Ma D.P.,
RA Nessler C.L., Thomas T.L.,
RT "Sequence and expression of a gene encoding an albumin storage
RT protein in sunflower."
RL Mol. Genet. 210:211-218(1987).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- PFM: IT IS POSSIBLE THAT THE 38 KDA PRECURSOR IS CLEAVED INTO TWO
CC POLYPEPTIDES THAT ARE APPROXIMATELY THE SAME SIZE. THE MATURE
CC PROTEIN IS COMPOSED OF A SINGLE POLYPEPTIDE CONTAINING ONE OR MORE
CC INTRA-MOLECULAR DISULFIDE LINKAGES.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: X06410; CAA29699.1; -
DR PIR: S01062; S01062.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereals_1tryp_1inh.
DR Pfam: PF00234; tryp_alpha_1tryp_1.
DR SMART: SM00439; AAI; 2.
KM Seed storage protein; Albumin; Signal; Multigene family.
FT SIGNAL 1 20
FT PROPEP 21 161 POTENTIAL.
FT CHAIN 162 295 2S SEED STORAGE PROTEIN.
FT SEQUENCE 295 AA; 34071 MW; 8958A106050142A1 CRC64;

Query Match 34.1%; Score 119.5; DB 1; Length 295;
Best Local Similarity 30.3%; Pred. No. 1.2e-06;
Matches 23; Conservative 17; Mismatches 23; Indels 13; Gaps 2;

QY 1 GERSLNG--CCDHLKOMQSCRCRGLRQAI-----EEQSQSGQLQGDVEAFRTA 47
DB 215 GQDQDRLQDQCCNCLQNVKRECHCAIQDVARRVRRKQFQDQDQDQDQDQDQDQDQ 274

QY 48 TAANLPSMCGVSPTEC 63
DB 275 VIONLPRQCDELEVOQC 290

RESULT 7
ID 2SSL_ARATH STANDARD; PRT; 164 AA.
AC P15457;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2S seed storage protein 1 precursor (2S albumin storage protein)
DE (NMW02-2S albumin 1).
GN AT2S1 OR AT4G27140 OR T24A18.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-73 AND 84-162.
RC STRAIN=CV. C24;
RA Kriebbers E., Herdies L., de Clercq A., Seurinck J., Leemans J.,
RA Van Damme J., Segura M., Gheysen G., Van Montagu M.,
RA Vandekerckhove J.;
RT "Determination of the processing sites of an Arabidopsis 2S albumin
RT and characterization of the complete gene family."
RL Plant Physiol. 87:859-866(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Conceicao A.D.S., Kriebbers E.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
RA Harris B., Ansoorge W., Brandt P., Griwell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermayer B., Macho R., Mueller M.,
RA Kreis M., Delseny M., Puldomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portelle D., Perez-Alonso M., Botry M., Bancroft I.,
RA Vos P., Hobeisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Wellens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Holtelegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., Van Staveren M., Dirksen W.,
RA Moollman P., Klein lankhorst R., Rose M., Hauf J., Koetler P.,
RA Bernieris S., Hempel S., Feldpausch M., Lambertz S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Kogers J., Cronin A., Quail M., Bryl-Alen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McWay K., Mayes R.,
RA Pertlett A., Rajandream M.A., Lyne M., Benes V., Rechin S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,
RA Neumann S., Agirion A., Vitale D., Liguori R., Piravandi E.,
RA Massenet O., Quidley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Cheifor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frisman D., Haase D., Lemcke K., Mews H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Clond J., Abbott A., Scott K., Johnson D.,

Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D.,
Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
Ma P., Zhong J., Preston R., Yil D., Shekher M., Matero A., Shah R.,
Swamy I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
Grat S., Shody N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
Chen E., Marra M., Marienssen R., McCombie W.R.;
"Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana." Nature 402:769-777(1999).
RL Nature 402:769-777(1999).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- MISCELLANEOUS: THIS IS THE MOST ABUNDANT ISOFORM OF 2S ALBUMIN IN
CC ARABIDOPSIS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M22032; AAA32743.1; -
DR EMBL: Z24745; CAA80870.1; -
DR EMBL: A13820; CAA01132.1; -
DR EMBL: AL035680; CAB38844.1; -
DR EMBL: AL161566; CAB79569.1; -
DR PIR: JAO161; NMM01.
DR PIR: S34676; S34676.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal-tryp_aml_inh.
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; tryp_alpha_aml; 1.
DR PRINTS: PR00496; Napin.
DR PRODOM: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Albumin; signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 37
FT CHAIN 38 73
FT PROPEP 74 83
FT CHAIN 84 162
FT PROPEP 163 164
SQ SEQUENCE 164 AA; 19014 MW; 2BF28CB474D9832B CRC64;

Query Match 32.0%; Score 112; DB 1; Length 164;
Best Local Similarity 35.8%; Pred. No. 4.9e-06;
Matches 24; Conservative 16; Mismatches 25; Indels 2; Gaps 2;

OY 1 OERSL-RGCGCHLKQMOGCRCEGLRQAIHQOOSGQLOGQVFAFRNALPSCGVS 59
DB 89 QEOQLPQOCNRLROEEDPCVPTLKQAKAVRLOGHQHQPQVRIYOTAKHLPVVCDIP 148
OY 60 PTE-CRF 65
DB 149 QVDVCPF 155

RESULT 8
PDB_WHEAT STANDARD; PRT; 148 AA.
AC 010464;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Puroindoline-B precursor.
OS Triticum aestivum (wheat).
OS

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CAPITOLE; TISSUE=Seed;
RX MEDLINE=94272013; PubMed=7516201;
RA Gautier M.-F., Aleman M.-F., Guirao A., Marion D., Joudrier P.;
RT "Triticum aestivum puroindolines, two basic cysteine-rich seed
RT proteins: cDNA sequence analysis and developmental gene expression.";
RL Plant Mol. Biol. 25:43-57(1994).
RN [2]
RP SEQUENCE OF 30-148.
RA Blochet J.E., Kaboulou A., Compont J.P., Marion D.;
RL (in) Bushuk W., Tkachuk R. (eds.);
RL Gluten proteins, pp.314-325, American Association of Cereal Chemists,
RL St. Paul MI (1991).
CC -1- FUNCTION: ACTS AS A MEMBRANOTOXIN, PROBABLY THROUGH ITS
CC ANTIMICROBIAL AND ANTIFUNGAL ACTIVITIES, CONTRIBUTING TO THE
CC DEFENSE MECHANISM OF THE PLANT AGAINST PREDATORS.
CC -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69912; CAA49537.1; -
DR HSSP: P07597; IITP.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal-tryp_aml_inh.
DR Pfam: PF00234; tryp_alpha_aml; 1.
DR SMART: SM00499; AAI; 1.
KW Membrane; Toxin; Antibiotic; Signal.
FT SIGNAL 1 19
FT PROPEP 20 29
FT CHAIN 30 148
FT DOMAIN 68 73
SQ SEQUENCE 148 AA; 16792 MW; 327904B4BEBC2C16 CRC64;

Query Match 30.9%; Score 108; DB 1; Length 148;
Best Local Similarity 35.9%; Pred. No. 1.3e-05;
Matches 23; Conservative 12; Mismatches 17; Indels 12; Gaps 3;

OY 8 CCDHLKQMOGCRCEGLRQAIHQOOSGQLOG-----QDVFAFRNALPSCGVSPT 61
DB 85 CCKOLSOIAPQCRDCSIRRVI-----QGRLGGLGIMRGEVFRKOLRAQSLPSKCMNG-A 138
OY 62 ECRF 65
DB 139 DCKF 142

RESULT 9
2SS2_CAPMA
ID 2SS2_CAPMA STANDARD; PRT; 155 AA.
AC P30233; O04774;
DT 01-APR-1993 (Rel. 25, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mabinlin II precursor (MAB II) (Sweet protein).
OS Capparis masakal (Mabinlang).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Capparis.
OX NCBI_TaxID=13395;
RN [1]
RP SEQUENCE FROM N.A.

```
RC TISSUE-Seed;
RX MEDLINE=97128796; PubMed=8973336;
RA Nirasawa S., Masuda Y., Nakaya K., Kurihara Y.;
RT "Cloning and sequencing of a cDNA encoding a heat-stable sweet
RL protein, mablin II.";
RN Gene 181:225-227(1996).
RM [2]
RP SEQUENCE OF 36-68 AND 83-154.
RC TISSUE-Seed;
RX MEDLINE=931145958; PubMed=8425538;
RA Liu X., Maeda S., Hu Z., Aiuchi T., Nakaya K., Kurihara Y.;
RT "Purification, complete amino acid sequence and structural
RL characterization of the heat-stable sweet protein, mablin II.";
RN Eur. J. Biochem. 211:281-287(1993).
RM [3]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE=94002261; PubMed=8399391;
RA Nirasawa S., Liu X., Nishino T., Kurihara Y.;
RT "Disulfide bridge structure of the heat-stable sweet protein mablin
RL II.";
RN Biochim. Biophys. Acta 1202:277-280(1993).
CC -1 FUNCTION: HEAT STABLE 2S SEED STORAGE PROTEIN HAVING SWEETNESS-
CC INDUCING ACTIVITY.
CC CC -1 SUBUNIT: HETERODIMER OF A SMALL A AND A LARGE B CHAIN LINKED BY
CC DISULFIDE BONDS.
CC -1 SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: D83997; BAA12204.1;
DR Medel: 14634; Capma:1175;14634.
DR InterPro: IPR003612; AAF.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_25; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00459; AAI; 1.
KW Seed storage protein; Albumin; Signal; Sweet-taste.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 35
FT CHAIN 36 68 MABLIN II, A CHAIN.
FT PROPEP 69 82
FT CHAIN 83 154 MABLIN II, B CHAIN.
FT PROPEP 155 155
FT DISULFID 40 103
FT DISULFID 53 92
FT DISULFID 93 141
FT DISULFID 105 149
FT MOD_RES 36 36 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 83 83 PYRROLIDONE CARBOXYLIC ACID.
FT CONFLICT 148 148 A->T (IN REF. 1).
FT CONFLICT 153 153 A->T (IN REF. 1).
SQ SEQUENCE 155 AA; 18089 MW; 72E8B5DEDC2D046A CRC64;
```

```

AC      P01090; 1986 (Rel. 01, Created)
AD      21-JUN-1986 (Rel. 01, last sequence update)
AE      01-MAR-1989 (Rel. 10, last annotation update)
AF      01-AUG-1992 (Rel. 23, last annotation update)
AG      Napin 2 precursor (1.7S seed storage protein).
AH      Brassica napus (Rape).
AI      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AJ      Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
AK      eurosids II; Brassicales; Brassicaceae; Brassica.
AL      NCBI_TaxID=3708;
AM      [1]
AN      SEQUENCE FROM N.A.
AO      MEDLINE=87308224; PubMed=3624251;
AP      Josefsson L.-G., Lemman M., Ericsson M.L., Rask L.;
AQ      "Structure of a gene encoding the 1.7 S storage protein, napin, from
AR      Brassica napus."
AS      J. Biol. Chem. 262:12196-12201(1987).
AT      [2]
AU      REVISIONS.
AV      Josefsson L.-G.;
AW      Submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.
AX      [3]
AY      SEQUENCE FROM N.A.
AZ      MEDLINE=87033665; PubMed=3771543;
BA      Ericsson M.L., Rosdin J., Lemman M., Glimelius K., Josefsson L.-G.,
BB      Rask L.;
BC      "Structure of the rapeseed 1.7 S storage protein, napin, and its
BD      precursor."
BE      J. Biol. Chem. 261:14576-14581(1986).
BF      [4]
BG      SEQUENCE FROM N.A.
BH      STRAIN=CV, TOWER;
BI      MEDLINE=84113267; PubMed=6689334;
BJ      Crouch M.L., Tenbarge K.M., Simon A.E., Fehr R.;
BK      "CNU clones for Brassica napus seed storage proteins: evidence from
BL      nucleotide sequence analysis that both subunits of napin are cleaved
BM      from a precursor polypeptide."
BN      J. Mol. Appl. Genet. 2:273-283(1983)
BO      [5]
BP      TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
BQ      ITS MATURATION.
BR      -1- SUBUNIT: THE MAJURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
BS      LINKED BY DISULFIDE BONDS.
BT      -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
BU      -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
BV      -----
BW      This SWISS-PROT entry is copyright. It is produced through a collaboration
BX      between the Swiss Institute of Bioinformatics and the EMBL outstation -
BY      the European Bioinformatics Institute. There are no restrictions on its
BZ      use by non-profit institutions as long as its content is in no way
CA      modified and this statement is not removed. Usage of and for commercial
CB      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CD      -----
CE      EMBL; K01545; AAA33006.1; -
CF      EMBL; J02586; AAA33297.1; -
CG      EMBL; J02798; AAA87348.1; -
CH      PIR; A01329; NMRP2.
CI      PIR; A29801; A29801.
CJ      PIR; A25997; A25997.
CK      InterPro; IPR003612; AAI.
CL      InterPro; IPR001768; cereal_1typ_aml_1nh.
CM      InterPro; IPR000617; Napin.
CN      Pfam; PF00234; 1typ_alpha_aml; 1.
CO      PRINTS; PR00496; NAPIN.
CP      PRODOM; PD002498; Napin; 1.
CQ      SMART; SM00499; AAI; 1.
CR      Seed storage protein; Signal; Multigene family.
CS      SIGNAL
CT      PROPEP      1      21
CU      FT          38
CV      FT CHAIN    39      74      SMALL CHAIN.
CW      PROPEP      75      94
CX      FT CHAIN    95      175      LARGE CHAIN.

```


FT CONFLICT 37 37 D -> N (IN REF. 4).
 FT CONFLICT 76 76 S -> N (IN REF. 4).
 SQ SEQUENCE 178 AA: 20104 MW: 734E561971B539FF CRC64:

Query Match 29.9%; Score 104.5; DB 1; Length 178;
 Best Local Similarity 38.2%; Pred. No. 3.8e-05;
 Matches 26; Conservative 12; Mismatches 23; Indels 7; Gaps 4;

OY 5 LRGCCDHLKQKOSQRC---EGIRQAIEQO-QSOGQLOG--QDVEAFRTANLPSMCGV 58
 DB 102 LQCCNELHQEPLCVCPITKASKAVKQOIQGQGGKLMVSRITYGTATHLPRVCNI 161
 OY 59 SPTF-CRF 65
 DB 162 PQSVYCPF 169

RESULT 11
 2SSL_PICGL STANDARD; PRT: 172 AA.

AC P26986; 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 25 seed storage-like protein precursor.
 OS Picea glauca (White spruce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
 OX NCBI_TaxID=3330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PG118;
 RA Newton C.H.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE 25 SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC

CC EMBL: X63193; CAA44875.1; -
 DR PIR: S18871; S18871.
 DR HSSP: P01085; IHSS.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal-tryp_amyl_inh.
 DR Pfam: PF00234; tryf_alpha_amyl_1.
 DR SMART: SM00499; AAI; 1.
 DR Seed storage protein; Albumin; Signal.
 FT SIGNAL 1 35
 FT CHAIN 36 172 2S SEED STORAGE-LIKE PROTEIN.
 FT SEQUENCE 172 AA: 20393 MW: 4C29E343B61003FB CRC64;

Query Match 29.7%; Score 104; DB 1; Length 172;
 Best Local Similarity 27.1%; Pred. No. 4.2e-05;
 Matches 23; Conservative 14; Mismatches 26; Indels 22; Gaps 2;
 OY 1 GRSRLRGCCDHLKQKOSQRC---EGIRQAIEQO-QSOGQLOG--QDVEAFRTANLPSMCGV 58
 DB 72 RQPSERCCCEELQKMSPOCRQAIOQMDQSLSYDFMDSQSDAPLNRRRREGGR 131
 OY 41 DVFEAFRTANLPSMCGV--SPTFEC 63
 DB 132 EEEEMERAAVLPNTCNVREPPRC 156

RESULT 12
 2SSB_BRANA

ID 2SSB_BRANA STANDARD; PRT: 178 AA.
 AC P27740;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Napin B precursor (1.7S seed storage protein).
 GN NABP.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SVALOFS KARAT 20516-K;
 RX MEDLINE=91231016; PubMed=2029903;
 RA Ericson M.L., Muren E., Gustavsson H.O., Josefsson L.G., Rask L.;
 RT "Analysis of the promoter region of napin genes from Brassica napus
 RT demonstrates binding of nuclear protein in vitro to a conserved
 RT sequence motif."
 RL Eur. J. Biochem. 197;741-746(1991).

CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPIINS ARE ONE OF THE
 CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
 CC ITS MATURATION.
 CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
 CC LINKED BY DISULFIDE BONDS.
 CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
 CC -1- SIMILARITY: BELONGS TO THE 25 SEED STORAGE ALBUMINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC

CC EMBL: X58142; CAA41150.1; -
 DR PIR: S15382; S15382.
 DR InterPro: IPR003612; AAI.
 DR InterPro: IPR001768; Cereal-tryp_amyl_inh.
 DR InterPro: IPR000617; Napin.
 DR Pfam: PF00234; tryf_alpha_amyl_1.
 DR PRINTS: PR00496; NABP.
 DR ProDom: PD002498; Napin; 1.
 DR SMART: SM00499; AAI; 1.
 DR Seed storage protein; Signal; Multigene family; Embryo.
 FT SIGNAL 1 21
 FT PROPEP 22 38
 FT CHAIN 39 74
 FT PROPEP 75 94
 FT CHAIN 95 178
 FT SEQUENCE 178 AA: 20114 MW: 96CE0ADB7CD966E9 CRC64;

Query Match 29.3%; Score 102.5; DB 1; Length 178;
 Best Local Similarity 38.2%; Pred. No. 6.5e-05;
 Matches 26; Conservative 12; Mismatches 23; Indels 7; Gaps 4;
 OY 5 LRGCCDHLKQKOSQRC---EGIRQAIEQO-QSOGQLOG--QDVEAFRTANLPSMCGV 58
 DB 102 LQCCNELHQEPLCVCPITKASKAVKQOIQGQGGKLMVSRITYGTATHLPRVCNI 161
 OY 59 SPTF-CRF 65
 DB 162 PQSVYCPF 169

RESULT 13
 ITRY_SINAR STANDARD; PRT: 130 AA.
 ID ITRY_SINAR
 AC P38057;
 DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin Inhibitor (TISA).
OS Sinapis arvensis (Charlock).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=29728;
RN [1]
RP SEQUENCE.
RC Tissue=Seed;
RA MEDLINE=94350545; PubMed=8070965;
RA Svendsen I.B., Nicolova D., Goshvay I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
chain trypsin inhibitor from the seeds of charlock (Sinapis arvensis
L.), a member of the napin protein family.";
RL Int. J. Pept. Protein Res. 43:425-430(1994).
CC -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10⁶ (6) M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Seed storage protein; Multigene family; Polymorphism.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R -> M.
FT VARIANT 53 57 MISSING (IN FORM II).
FT VARIANT 73 73 A -> S.
FT VARIANT 77 77 K -> R.
FT VARIANT 81 81 Q -> R.
FT VARIANT 87 87 Q -> Q.
FT VARIANT 89 89 Q -> H.
FT VARIANT 91 91 G -> Q.
FT VARIANT 97 97 E -> M.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
SQ SEQUENCE 130 AA; 14682 MW; EC02F4B26D180DF2 CRC64;

Query Match 29.1%; Score 102; DB 1; Length 130;
Best Local Similarity 32.9%; Pred. No. 5.4e-05;
Matches 24; Conservative 15; Mismatches 22; Indels 12; Gaps 4;

OY 5 LRCCDHLKQMSQRC---EGLRQAIHQ-QSQGLOGQ-----DVFEAFRTAANLP 53
DB 50 LQCCNEHQBEPICVPTLKGAAKAVKQIQQGQGGQGLQHHEIRIRYQATATLP 109
OY 54 SMCQSPTE-CRF 65
DB 110 KVCNIPQVQCPF 122

RESULT 14
ALL_BRAJU STANDARD; PRT; 129 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Allergen Bra j 1-E, small and large chains (Bra j 1).
OS Brassica juncea (Leaf mustard) (Indian mustard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3707;
RN [1]
RP SEQUENCE.
RC Tissue=Seed;
RA MEDLINE=93356721; PubMed=7688955;
RA Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Arias L.,
RA Lopez-Otin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope.";
RL Biochem. J. 293:625-632(1993)
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY TWO DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR Pfam: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIN.
DR ProDom: PD002498; Napin; 1.
DR SMART: SM00499; AAI; 1.
KW Allergen; Seed storage protein.
FT CHAIN 1 37 SMALL CHAIN.
FT NON_CONS 37 38
FT CHAIN 38 129 LARGE CHAIN.
FT VARIANT 6 6 F -> I.
FT VARIANT 20 20 R -> K.
SQ SEQUENCE 129 AA; 14644 MW; D6F2B03F62B08F8 CRC64;

Query Match 28.7%; Score 100.5; DB 1; Length 129;
Best Local Similarity 33.8%; Pred. No. 7.9e-05;
Matches 22; Conservative 15; Mismatches 17; Indels 11; Gaps 3;

OY 5 LRCCDHLKQMSQRC---EGLRQAIHQ-QSQGLOGQ-----DVFEAFRTAANLP 53
DB 49 LQCCNEHQBEPICVPTLKGAAKAVKQIQQGQGGQGLQHHEIRIRYQATATLP 108
OY 54 SMCQSV 58
DB 109 RVCNI 113

RESULT 15
2SSE_BRAJU STANDARD; PRT; 186 AA.
AC P09893;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Napin embryo specific precursor (1.7S seed storage protein).
OS Brassica napus (Rape)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87308225; PubMed=3040733;
RA Scofield S.R., Crouch M.L.;
RT "Nucleotide sequence of a member of the napin storage protein family
from Brassica napus.";
RL J. Biol. Chem. 262:12202-12208(1987).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -1- DEVELOPMENTAL STAGE: EMBRYO.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on how
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; J02782; AAA33007.1; -.	
DR	PIR; A29802; A29802.	
DR	InterPro; IPR003612; AAI.	
DR	InterPro; IPR01768; Cereal_ttyp_aml1nh.	
DR	InterPro; IPR006617; Napin.	
DR	Pfam; PF00234; ttyp_alpha_aml1.1.	
DR	PRINTS; PR00496; NAPIN.	
DR	ProDom; PD002498; Napin.1.	
DR	SMART; SM00499; AAI.1.	
KW	Seed storage protein; Signal; Multigene family.	
FT	SIGNAL	1..21
FT	PROPEP	22..38
FT	CHAIN	39..76
FT	PROPEP	77..97
FT	CHAIN	98..186
SO	SEQUENCE	186 AA; 21013 MW; 9CAE65D84B160AB3 CRC64

Query Match	28.1%	Score 98.5	DB 1	Length 186
Best Local Similarity	32.9%	Pred. No. 0.00019		
Matches 23, Conservative 12, Mismatches 22, Indels 13, Gaps 4				

```

QY      8  CCDHKKQMOQSCRC-----EGLRQAI EEOQSOLOGOD---VEAPRTTANLPSC 56
      11  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      108 CCNELHDEEPRLCYCPITLKGSKAVRQGVRRQGG-QMGQGMQVYSRVYQTATHLPRVC 166
QY      57 GVSFPE-CRF 65
      167 NIROYSLCPF 176
Db

```

Search completed: September 9, 2002, 12:46:06
Job time: 308 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2002, 12:45:37 : Search time 77.42 seconds
(without alignments)
145.242 Million cell updates/sec

Title: US-09-913-351-5

Perfect score: 350
Sequence: 1 QERSLRGCCDHLKQMSOCK.....FRTANLPMSGVSPTRECR 65

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	347	99.1	65	10 Q9S872
2	165	47.1	139	10 P93198
3	158	45.1	153	10 Q9AUD1
4	150.5	43.0	64	10 Q9AVK2
5	132.5	37.9	146	10 Q9LRC2
6	130.5	37.3	165	10 Q9FH31
7	130	37.1	141	10 Q9S649
8	127	36.3	66	10 Q9S870
9	124	35.4	148	10 Q9XHP1
10	118.5	33.9	161	10 Q9L410
11	113.5	32.4	60	10 Q93YG0
12	113.5	32.4	165	10 Q40850
13	112.5	32.1	139	10 Q39787
14	112.5	32.1	139	10 Q39795
15	110.5	31.6	323	10 Q39928
16	109	31.1	148	10 Q9M4E4

17	108.5	31.0	148	10 Q9M4D9	Q9M4D9
18	107.5	30.7	167	10 Q81411	Q81411
19	107	30.6	148	10 Q9SC07	Q9SC07
20	106	30.3	147	10 Q9FS19	Q9FS19
21	105	30.0	147	10 Q9AWH0	Q9AWH0
22	105	30.0	148	10 Q9AMG9	Q9AMG9
23	104.5	29.9	86	10 Q9S9F0	Q9S9F0
24	104.5	29.9	178	10 Q42413	Q42413
25	104.5	29.9	178	10 Q39344	Q39344
26	104	29.7	162	10 Q64929	Q64929
27	103.5	29.6	173	10 Q9ZRH1	Q9ZRH1
28	103	29.4	148	10 Q9AQX9	Q9AQX9
29	102.5	29.3	173	10 Q81412	Q81412
30	101.5	29.0	155	10 Q41167	Q41167
31	101.5	29.0	178	10 Q42490	Q42490
32	101	28.9	156	10 Q94180	Q94180
33	101	28.9	190	10 Q40995	Q40995
34	100	28.6	148	10 Q9M4E1	Q9M4E1
35	99.5	28.4	162	10 Q64932	Q64932
36	99	28.3	91	10 Q9S9E6	Q9S9E6
37	99	28.3	91	10 Q9S9E5	Q9S9E5
38	99	28.3	148	10 Q9AVP9	Q9AVP9
39	99	28.3	162	10 Q40998	Q40998
40	98.5	28.1	88	10 Q9S9E9	Q9S9E9
41	98.5	28.1	164	10 Q9FV75	Q9FV75
42	98	28.0	152	10 Q41168	Q41168
43	97.5	27.9	88	10 Q9S9E8	Q9S9E8
44	97	27.7	172	10 Q40997	Q40997
45	97	27.7	174	10 Q41169	Q41169

ALIGNMENTS

RESULT 1
Q9S872 PRELIMINARY: PRT: 65 AA.
ID Q9S872:
AC Q9S872:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 7.3 KDA NAPIIN-LIKE PROTEIN LARGE CHAIN (FRAGMENT).
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE.
RX MEDLINE=97135090; PubMed=8980648;
RA Neumann G.M., Condon R., Polya G.M.;
RT "Purification and sequencing of napin-like protein small and large
RT chains from Momordica charantia and Ricinus communis seeds and
RT determination of sites phosphorylated by plant Ca(2+)-dependent
RT protein kinase.";
RL Biochim. Biophys. Acta 1298:223-240(1996).
DR InterPro: IPR003612; AAI.
DR InterPro: IPR000617; Napin.
DR PRINTS: PR00496; NAPIN.
DR SMART: SM00499; AAI: 1.
SQ SEQUENCE 65 AA; 7350 MW; CB9264C4C917F498 CRC64;

Query Match 99.1%; Score 347; DB 10; Length 65;
Best Local Similarity 98.5%; Pred. No. 3.7e-35;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QERSLRGCCDHLKQMSOCKRCGLQALIEQDSQGLQSGDYFEAFRTANLPMSGVSPTRECR 60
:|||||
DB 1 EERSLRGCCDHLKQMSOCKRCGLQALIEQDSQGLQSGDYFEAFRTANLPMSGVSPTRECR 60
QY 61 TECRF 65
|||||

QY 8 CCHLHKOMOSCREGLRQAIEOOSO-----GOLGODVFEARPTAANI.PSMCGVSPT E 62
|| :: : |||::||| | | :: : | : | : |
Db 73 CCQGLEKMDQCRCROGLRHATMQMOOMOQOMSCKOMRI MQ--KVTKNIMSECEMEPR 130

DR InterPro: IPR001768; Cereal_1-trypanin
DR InterPro: IPR000617; Napin.
DR Pfam: PF00234; 1-trypanin; 2.

DR SMART; SM00499; AAI; 2.
 KW Signal.
 FT NON_TER 1 1
 FT SIGNAL <1 19
 FT CHAIN 156 323
 SQ SEQUENCE 323 AA; 38282 MW; A15378AFAD52D6B3 CRC64;

Query Match 31.6%; Score 110.5; DB 10; Length 323;
 Best Local Similarity 30.9%; Pred. No. 1.5e-05;
 Matches 21; Conservative 15; Mismatches 27; Indels 5; Gaps 1;

QY 1 QERSLRGCCDHLKQWQSGRC-----GLRQALIEDQSQGLOGDPVFEAFRTANLPSM 55
 DB 252 EGGGLQCCCNELQNVRRRCQCAIKREVQGRMRQOQOORRQYGGQOTQTVERTLENLPNQ 311
 QY 56 CGVSPTEC 63
 DB 312 CDLDVQQC 319

Search completed: September 9, 2002, 12:45:37
 Job Time: 294 Sec